4-2012

Semiparametric GMM Estimation of Spatial Autoregressive Models

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

DOI: https://doi.org/10.1016/j.jeconom.2011.09.034

Published in Journal of Econometrics
Volume 167, Issue 2, April 2012, Pages 543-560
https://doi.org/10.1016/j.jeconom.2011.09.034

Semiparametric GMM estimation of spatial autoregressive models[☆]

Liangjun Su
School of Economics, Singapore Management University, Singapore

A B S T R A C T

We propose semiparametric GMM estimation of semiparametric spatial autoregressive (SAR) models under weak moment conditions. In comparison with the quasi-maximum-likelihood-based semiparametric estimator of Su and Jin (2010), we allow for both heteroscedasticity and spatial dependence in the error terms. We derive the limiting distributions of our estimators for both the parametric and nonparametric components in the model and demonstrate the estimator of the parametric component has the usual $\sqrt{n}$-asymptotics. When the error term also follows an SAR process, we propose an estimator for the parameter in the SAR error process and derive the joint asymptotic distribution for both spatial parameters. Consistent estimates for the asymptotic variance-covariance matrices of both the parametric and nonparametric components are provided. Monte Carlo simulations indicate that our estimators perform well in finite samples.

Keywords:
Generalized method of moments
Local instruments
Nonlinearity
Semiparametrics
Spatial autoregression

1. Introduction

Nonlinearity is important in spatial dependence models. As was argued in Paelinck and Klaassen (1979, pp. 6–9), econometric relations in space result more often than not in highly nonlinear specifications. Despite this observation, most theoretical studies on spatial autoregressive (SAR) models ignore potential nonlinear functional forms. A few exceptions include van Gastel and Paelinck (1995), Baltagi and Li (2001), Pace et al. (2004), and Yang et al. (2006), who have considered flexible functional forms to account for certain forms of nonlinearities. On the application side, recent researchers have started addressing the importance of nonparametric modeling in spatial econometrics. For example, in modeling hedonic housing prices, Gress (2004) introduced two semiparametric spatial autocorrelation models and compare them with a variety of competing parametric spatial models, in which the exogenous regressors include house size, house age, latitude, longitude, and some dummy variables for the zip code area where the house resides. He found that the semiparametric models offer more accurate and stable estimates of the regression parameters and better out-of-sample predictions than do the alternative parametric models. Basile and Gress (2005) proposed a semiparametric spatial auto-covariance specification of the growth model for the European economy, where the dependent variable is the average per capita GDP growth rate in the period 1988–2000 and the exogenous regressors include the initial per capita GDP, average proportion of real physical investment to real GDP, average growth rate of the population, and average unemployment rate. They found that nonlinearities are important in regional growth in Europe even when the spatial dependence is controlled for. As a result, assuming a common linear relationship between economic growth and inputs is misleading.

Recently, Su and Jin (2010) propose a quasi-maximum-likelihood-based estimator of partially linear spatial autoregressive models and demonstrate that the rates of consistency for the finite-dimensional parameters in the model depend on some general features of the spatial weight matrix. Unfortunately, like the quasi-maximum likelihood estimator (QMLE) of Lee (2004) in the parametric setup, their estimator does not have a closed form expression and thus is not easy to implement in practice.

In this paper we propose semiparametric GMM (SPGMM hereafter) estimation of semiparametric SAR models where the spatial lag effect (endogenous variable) enters the model linearly and the exogenous variables enter the model nonparametrically. Based on some moment conditions implied by the model, we propose a two-stage estimation strategy for both the one-dimensional spatial parameter and the nonparametric component and term the resulting estimators as SPGMM estimators. In the first stage, we treat the spatial parameter as if it were known, and use some local instruments to estimate the nonparametric component locally as a function of the spatial parameter. In the

[☆] The author is grateful to the guest editors and two anonymous referees for their many constructive comments on an early version of this paper. The author gratefully acknowledges the financial support from the NSFC under grant numbers 70501001 and 70601001.

E-mail address: ljsu@smu.edu.sg.

second stage we use the global instruments to estimate the spatial parameter by profiling out the nonparametric component, and recover the estimate of the nonparametric component. It is worth mentioning that the idea of nonparametric profiling is not new in the literature; see Su and Ullah (2006) and Henderson et al. (2008) for recent applications in econometrics. Nevertheless, to the best of our knowledge, this paper is the first to apply the idea of nonparametrically profiling out a nonparametric function of the exogenous regressor before applying instrumental variables to estimate the coefficient of the endogenous regressor. As a referee kindly points out, this idea seems quite general and can be applied to estimate various semiparametric conditional mean or quantile models with endogeneity in which the endogenous variables enter the model parametrically (linear or nonlinear) while the exogenous variables enter nonparametrically.

In comparison with the estimation strategy of Su and Jin (2010), our SPGMM approach has several advantages. First, our estimation is based upon some moment conditions instead of a quasi-likelihood function. As a result, we can have an analytic form for our estimator, and it is easy to implement in practice. Second, we can obtain the usual parametric consistency rate for our estimator of the parametric component in the model whereas the consistency rate of Su and Jin's (2010) estimator depends on some general features of the spatial weight matrix. Third, unlike Su and Jin (2010) who only consider homoscedastic and independent errors, we allow the stochastic error terms to exhibit both heteroscedasticity of unknown form and a certain form of spatial dependence. Fourth, we allow for both continuous and discrete exogenous regressors in the model whereas Su and Jin (2010) consider only continuous regressors in their nonparametric component. Fifth, our method can be easily extended to semiparametric spatial panel data autoregressive models or applied to other types of semiparametric models with endogeneity.

The paper is structured as follows. In Section 2 we introduce the semiparametric SAR model and the semiparametric GMM approach to estimate the finite and infinite dimensional parameters in the model. In Section 3 we first make some basic assumptions underlying our analysis and then study the asymptotic properties of the estimators for both the parametric and nonparametric parts. We explore the estimation of the spatial parameter in the spatial error process and derive the joint asymptotic distribution of both spatial parameters in Section 4. Section 5 discusses consistent estimation of the asymptotic variance-covariance matrices. We conduct a small set of Monte Carlo simulations to check the finite-sample performance of the proposed estimators in Section 6. Final remarks are contained in Section 7. All technical details are relegated to the appendices.

Like Kelejian and Prucha (2001), we adopt the following notation and conventions. For a matrix A_n, we denote its norm as $\|A_n\| = \left[\operatorname{tr}\left(A_nA_n'\right)\right]^{1/2}$, its (i,j)th element as $a_{n,ij}$, and its minimum eigenvalue as $\lambda_{\min}(A_n)$ when A_n is a square matrix. For a vector a_n we use $a_{n,i}$ to denote its ith element and $\operatorname{diag}(a_n)$ a diagonal matrix with $a_{n,i}$ as its (i,i)th element. An analogous convention is adopted for matrices and vectors that do not depend on the index n, where n is frequently suppressed. We say A_n is uniformly bounded in absolute value if $\sup_{1\le i\le n, 1\le j\le n}\left|a_{n,ij}\right| < c$ for some $c < \infty$. Following Lee (2002), we say that the row (resp. column) sums of A_n are uniformly bounded in absolute value if $\sup_{1\le i\le n, n\ge 1}\sum_{j=1}^{n}\left|a_{n,ij}\right| \le c_a < \infty$ (resp. $\sup_{1\le j\le n, n\ge 1}\sum_{i=1}^{n}\left|a_{n,ij}\right| \le c_a < \infty$). Similarly, we say that the row (resp. column) sums of A_n are uniformly bounded in absolute value for sufficiently large n if $\sup_{1\le i\le n, n\ge N_1}\sum_{j=1}^{n}\left|a_{n,ij}\right| \le c_a < \infty$ (resp. $\sup_{1\le j\le n, n\ge N_1}\sum_{i=1}^{n}\left|a_{n,ij}\right| \le c_a < \infty$) for some large integer N_1. For $p_c \times 1$ vectors $a \equiv (a_1, \ldots, a_{p_c})'$ and $b \equiv (b_1, \ldots, b_{p_c})'$, define $a/b = (a_1/b_1, \ldots, a_{p_c}/b_{p_c})'$. Let $\mathbf{0}_{d_1\times d_2}$ denote a $d_1 \times d_2$ matrix of zeros. Let $\odot$ and $\otimes$ denote the Hadamard and Kronecker products, respectively.

2. Model and semiparametric GMM estimation

2.1. Model and moment conditions

Consider the following semiparametric spatial autoregressive model:

$$Y_n = \mathbf{m}(X_n) + \rho_n^0 W_{1n} Y_n + U_n, \tag{2.1}$$

where $X_n \equiv (x_{n,1}, \ldots, x_{n,n})'$ is an $n \times p$ matrix of fixed regressors that do not contain the constant term, W_{1n} is a prespecified constant $n \times n$ spatial weight matrix, $\mathbf{m}(X_n) \equiv \left(m(x_{n,1}), \ldots, m(x_{n,n})\right)'$, $m(\cdot)$ is an unknown function defined on $\mathbb{R}^p$, and $U_n \equiv (u_{n,1}, \ldots, u_{n,n})'$ is an n-dimensional vector of zero mean random variables that are not necessarily identically distributed and may also exhibit spatial dependence structure such as spatial autoregressive or spatial moving average (SMA) forms. When U_n also exhibits an SAR form, we will call the model in (2.1) as a semiparametric SARAR model. When U_n exhibits an SMA form, we will call the model in (2.1) as a semiparametric SARMA model.

If $W_{1n}Y_n$ were not endogenous, we could extend the procedure of Robinson (1988) to our framework and estimate both the parametric component and nonparametric component in (2.1). Nevertheless, since $W_{1n}Y_n$ is endogenously generated here, one can show that the estimator of Robinson (1988) is generally inconsistent. (An exception occurs when the elements of the spatial weight matrix W_{1n} are uniformly of the order $o(n^{-1/2})$.)

Let $\overline{Y}_n \equiv W_{1n}Y_n$ and denote its ith element as $\overline{y}_{n,i}$. When $\rho_n^0 \neq 0$, we can assume that there exists a $q \times 1$ vector of nonstochastic instruments $z_{n,i}$ for $\widetilde{x}_{n,i} \equiv \left(x_{n,i}', \overline{y}_{n,i}\right)'$ such that we have the following orthogonality condition

$$E\left(z_{n,i}u_{n,i}\right) = 0. \tag{2.2}$$

For example, $z_{n,i} = \left(1, x_{n,i}', \overline{x}_{n,i}'\right)'$, where $\overline{x}_{n,i}$ is the ith row of $\overline{X}_n \equiv W_{1n}X_n$. In the following, we assume that $z_{n,i}$ contains 1. We will propose semiparametric GMM estimation of both ρ_n^0 and $m(\cdot)$ based upon (2.2).

2.2. Semiparametric GMM estimation

The moment condition in (2.2) implies that[1]

$$E\left\{z_{n,i}\left[y_{n,i} - \rho_n^0\overline{y}_{n,i} - m\left(x_{n,i}\right)\right]\right\} = 0. \tag{2.3}$$

Clearly, (2.3) provides moment restrictions and can lead to an estimation approach similar to the generalized method of moments (GMM) procedure of Hansen (1982) for parametric models. Since the functional form $m(\cdot)$ is unknown, one can approximate it with a sieve and estimate ρ_n^0 and the sieve parameters jointly. Such an approach was taken by Ai and Chen (2003) who show that the sieve estimator of the nonparametric component (m here) is consistent with a rate faster than $n^{-1/4}$ under certain metric, and the estimator of the parametric component (ρ_n^0 here) is $\sqrt{n}$-consistent and asymptotically normally distributed. It is worth mentioning that the data are assumed to be independent and identically distributed (i.i.d.) in Ai and Chen (2003) so that their theory cannot be applied directly to our framework. In addition, the asymptotic distribution of the estimator of the nonparametric component is also of our main interest, which is unfortunately not addressed in Ai and Chen (2003).

[1] Alternatively, we can impose the following conditional moment restrictions: $E\{y_{n,i} - \rho_n^0\overline{y}_{n,i} - m\left(x_{n,i}\right)|z_{n,i}\} = 0$, where $z_{n,i}$ contains $x_{n,i}$ and is treated as random. In this case, we have $E\{\mathbf{b}\left(z_{n,i}\right)[y_{n,i} - \rho_n^0\overline{y}_{n,i} - m\left(x_{n,i}\right)]\}$ for any vector function $\mathbf{b}\left(z_{n,i}\right)$.

In this paper we propose to approximate $m(\cdot)$ by using the local linear fitting method of Fan (1992) and Fan and Gijbels (1996). Fan and Gijbels (1996) have documented the advantages of local linear method or, more generally, the local polynomial method, over the conventional local constant (Nadaraya–Watson) method or the sieve/series method. In particular, when the support of $x_{n,i}$ is compact, the local linear/polynomial method automatically adjusts to boundary points so that it is not subject to the "boundary bias" problem associated with the local constant method. In comparison with the sieve method, one can readily establish asymptotic normality for the local linear/polynomial estimator. In the following, we will focus on the local linear method for the sake of notational simplicity. Nevertheless, as remarked after Assumption 4, if $x_{n,i}$ contains at least four continuous regressors, one will need to use a higher order local polynomial as in Masry (1996).

To allow for both continuous and discrete regressors in $x_{n,i}$, write $x_{n,i} = (x_{n,i}^{c'}, x_{n,i}^{d'})'$ where $x_{n,i}^c$ denotes a $p_c \times 1$ vector of continuous regressors in $x_{n,i}$ and $x_{n,i}^d$ denotes a $p_d \times 1$ vector of remaining discrete regressors with $p_d = p - p_c$. We assume that some of the discrete regressors have a natural ordering, examples of which would include environmental conditions (excellent, good, poor) or preference ordering (like, indifference, dislike) etc. Let $\overrightarrow{x}_{n,i}^{d}$ denote a $p_1 \times 1$ vector (say, the first p_1 components of $x_{n,i}^d$, $0 \le p_1 \le p_d$) of discrete regressors that have a natural ordering. Let $\widetilde{x}_{n,i}^d$ denote the remaining $p_2 = p_d - p_1$ discrete regressors that do not have a natural ordering. We use $x_{n,is}^c$ and $x_{n,is}^d$ to denote the sth element of $x_{n,i}^c$ and $x_{n,i}^d$, respectively ($s = 1, \ldots, p_c$ or p_d).

For the continuous exogenous regressor, we choose a product kernel function $Q(\cdot)$ of $q(\cdot)$ and a vector of smoothing parameters $h = (h_1, \ldots, h_{p_c})'$. Let $Q_{h,i}(x^c) = \Pi_{s=1}^{p_c} h_s^{-1} q\left(\left(x_{n,is}^c - x_s^c\right)/h_s\right)$ and

$$Q_{h,ij} \equiv Q_{h,i}\left(x_{n,j}^c\right) = \prod_{s=1}^{p_c} h_s^{-1} q\left(\left(x_{n,is}^c - x_{n,js}^c\right)/h_s\right). \tag{2.4}$$

For the unordered discrete regressor, we follow Racine and Li (2004) and Li and Racine (2007) and use a variation of the kernel function of Aitchison and Aitken (1976):

$$\widetilde{l}\left(\widetilde{x}_{n,is}^d, \widetilde{x}_{n,js}^d, \lambda_s\right) = \begin{cases} 1 & \text{if } \widetilde{x}_{n,is}^d = \widetilde{x}_{n,js}^d \\ \lambda_s & \text{otherwise,} \end{cases} \tag{2.5}$$

where $\lambda_s \in [0, 1]$ is the smoothing parameter. In the special case where $\lambda_s = 0$, $\widetilde{l}(\cdot,\cdot,\cdot)$ reduces to the usual indicator function as used in the nonparametric frequency approach. Similarly, $\lambda_s = 1$ leads to a uniform weight function, in which case, the $\widetilde{x}_{n,is}^d$ regressor will be completely smoothed out in the sense that it will not affect the nonparametric estimation result. For the ordered discrete regressors, we assume that for $s = 1, \ldots, p_1$, $\overrightarrow{x}_{n,is}^{d}$ take only integer values $0, 1, 2, \ldots, c_s$, where $1 \le c_s < \infty$. We follow Racine and Li (2004) and use the following kernel

$$\overrightarrow{l}\left(\overrightarrow{x}_{n,is}^{d}, \overrightarrow{x}_{n,js}^{d}, \lambda_s\right) = \begin{cases} 1 & \text{if } \overrightarrow{x}_{n,is}^{d} = \overrightarrow{x}_{n,js}^{d} \\ \lambda_s^{\left|\overrightarrow{x}_{n,is}^{d} - \overrightarrow{x}_{n,js}^{d}\right|} & \text{otherwise.} \end{cases} \tag{2.6}$$

Again, choosing $\lambda_s = 0$ or 1 leads to similar remarks as above.

Combining (2.5) and (2.6), we obtain the product kernel function for the discrete regressors:

$$L_{\lambda,ij} \equiv L_{\lambda,i}\left(x_{n,j}^d\right) \equiv \left[\prod_{s=1}^{p_1} \lambda_s^{\left|\overrightarrow{x}_{n,is}^{d} - \overrightarrow{x}_{n,js}^{d}\right|}\right]\left[\prod_{s=1}^{p_2} \lambda_{s+p_1}^{1-\mathbf{1}\left(\widetilde{x}_{n,is}^d = \widetilde{x}_{n,js}^d\right)}\right], \tag{2.7}$$

where $\lambda = (\lambda_1, \ldots, \lambda_{p_d})'$, and $\mathbf{1}(A) = 1$ if A holds and 0 otherwise. Combining (2.4) and (2.7), we obtain the product kernel function for all the exogenous regressors:

$$K_{h\lambda,ij} \equiv K_{h\lambda,i}\left(x_{n,j}\right) = Q_{h,i}\left(x_{n,j}^c\right) L_{\lambda,i}\left(x_{n,j}^d\right). \tag{2.8}$$

Now, fix a point $x_{n,j} = (x_{n,j}^{c'}, x_{n,j}^{d'})'$. It follows from the first order Taylor expansion that

$$m\left(x_{n,i}\right) \approx m\left(x_{n,j}\right) + \dot{m}\left(x_{n,j}\right)'\left(x_{n,i}^c - x_{n,j}^c\right) \tag{2.9}$$

for any $x_{n,i}^c$ in the neighborhood of $x_{n,j}^c$ and $x_{n,i}^d = x_{n,j}^d$, where $\dot{m}(x) = \partial m(x)/\partial x^c$, i.e., the derivative is only taken with respect to the continuous component x^c of $x = (x^{c\prime}, x^{d\prime})'$. So (2.3) can be approximated as follows

$$E\left\{z_{n,i}\left[y_{n,i} - \rho_n^0 \bar{y}_{n,i} - m(x) - \dot{m}(x)'\left(x_{n,i}^c - x^c\right)\right]\right\} \approx 0, \tag{2.10}$$

where $x_{n,i}^c$ is close to x^c and $x_{n,i}^d = x^d$. If the above relationship held exactly, we could follow the approach of Newey (1990) to construct optimal instruments for the efficient estimation of both ρ_n^0 and $(m(x), \dot{m}(x)')$. Alternatively one could follow Ai and Chen (2003) and construct optimal instruments for the efficient estimation of ρ_n^0 based upon (2.3). In either case, difficulty arises here due to the approximation nature of the relationship in (2.10), the non-i.i.d. observations, potential heteroscedasticity of unknown form, and potential spatial dependence in the disturbances. For this reason, we will focus on a convenient choice of $z_{n,i}$ and leave the optimality issue as an open question. Furthermore, due to the local nature of the approximation in (2.10), we will allow the instruments $z_{n,i}$ to be *locally* dependent on the point of approximation (x) and certain parameters used in the approximation when we consider profiling out the nonparametric component.

Noting that the unknown parameters in (2.10) include both the global parameter ρ_n^0 and the nonparametric local parameter vector $(m(x), \dot{m}(x))$, we now propose a two-step procedure to estimate these unknown parameters by profiling out the nonparametric component first.

In the first step, we treat ρ_n^0 as if it were known in (2.10) and consider the estimation of $m(x)$ and $\dot{m}(x)$ as a function of ρ_n. At the sample level, the orthogonality condition in (2.10) implies the following locally weighted orthogonality conditions

$$n^{-1}\sum_{i=1}^{n} \mathbf{z}_{h,ij}\left\{y_{n,i} - \rho_n \bar{y}_{n,i} - \boldsymbol{\tau}_{h,ij}' \mathbf{M}_h\left(x_{n,j}\right)\right\} K_{h\lambda,ij} = 0, \tag{2.11}$$

where $\mathbf{M}_h(x)$ is a $(p_c + 1) \times 1$ vector of parameters whose true value corresponds to $(m(x), (h \odot \dot{m}(x))')'$, $\boldsymbol{\tau}_{h,ij} \equiv \boldsymbol{\tau}_{h,i}\left(x_{n,j}\right) = (1, ((x_{n,i}^c - x_{n,j}^c)/h)')'$. Here we allow the "*local instruments*" $\mathbf{z}_{h,ij}$ to depend on the smoothing parameter h and the particular point $x_{n,j}$ at which we approximate the function $m(\cdot)$. Motivated by the idea of local linear fitting, we can follow Cai and Li (2008) and choose $\mathbf{z}_{h,ij}$ simply as

$$\mathbf{z}_{h,ij} \equiv \mathbf{z}_{h,i}\left(x_{n,j}^c\right) = \begin{pmatrix} z_{n,i}^{(1)} \\ z_{n,i}^{(1)} \otimes \left(\left(x_{n,i}^c - x_{n,j}^c\right)/h\right) \end{pmatrix}, \tag{2.12}$$

where $\otimes$ is the Kronecker product and $z_{n,i}^{(1)}$ is a subset of $z_{n,i}$. In an extreme case, one can take $z_{n,i}^{(1)} = z_{n,i}$. In the other extreme case, it is fine to take $z_{n,i}^{(1)} = 1$ and then $\mathbf{z}_{h,ij} = \boldsymbol{\tau}_{h,ij}$, which results in the local linear profile estimator of $\mathbf{M}_h\left(x_{n,j}\right)$ by regressing $y_{n,i} - \rho_n \bar{y}_{n,i}$ on $x_{n,i}^c$ and treating ρ_n as if it were known. In either case, it is sufficient to identify $\mathbf{M}_h$ in (2.11) for any given ρ_n and point $x_{n,j}$. Noting that unless the dimension of $\mathbf{z}_{h,ij}$ is the same as that of $\boldsymbol{\tau}_{h,ij}$, the number of equations in (2.11) is greater than the number of parameters $(p_c + 1)$ for any fixed ρ_n and $x_{n,j}$, so that the model is overidentified and we may not have a unique $\mathbf{M}_h$ satisfying (2.11). To ensure a unique solution, we premultiply (2.11) by $\mathbf{A}_{n,h\lambda}\left(x_{n,j}\right)' \equiv n^{-1}\sum_{i=1}^{n} \boldsymbol{\tau}_{h,ij}\mathbf{z}_{h,ij}' K_{h\lambda,ij}$ to obtain

$$\mathbf{A}_{n,h\lambda}\left(x_{n,j}\right)' n^{-1}\sum_{i=1}^{n}\mathbf{z}_{h,ij}\left\{y_{n,i}-\rho_n\bar{y}_{n,i}-\boldsymbol{\tau}'_{h,ij}\mathbf{M}_h\left(x_{n,j}\right)\right\}K_{h\lambda,ij}=0. \tag{2.13}$$

Solving the above equation for $\mathbf{M}_h$ yields the following solution

$$\begin{aligned}\overline{\mathbf{M}}_{\rho_n,h\lambda}\left(x_{n,j}\right) &= \left(\mathbf{A}_{n,h\lambda}\left(x_{n,j}\right)'\mathbf{A}_{n,h\lambda}\left(x_{n,j}\right)\right)^{-1}\\ &\quad\times\mathbf{A}_{n,h\lambda}\left(x_{n,j}\right)'\mathbf{B}_{n\rho_n,h\lambda}\left(x_{n,j}\right)\end{aligned}$$

where $\mathbf{B}_{n\rho_n,h\lambda}\left(x_{n,j}\right) = n^{-1}\sum_{i=1}^{n}K_{h\lambda,ij}\mathbf{z}_{h,ij}\left(y_{n,i}-\rho_n\bar{y}_{n,i}\right)$. In particular, the estimator of $m\left(x_{n,j}\right)$ is given by

$$\begin{aligned}\overline{m}_{\rho_n,h\lambda}\left(x_{n,j}\right) &= e_1'\overline{\mathbf{M}}_{\rho_n,h\lambda}\left(x_{n,j}\right)\\ &= \mathbf{s}_{h\lambda}\left(x_{n,j}\right)'\left(Y_n-\rho_nW_{1n}Y_n\right),\end{aligned}$$

where $e_1\equiv(1,0,\ldots,0)'$ is a (p_c+1)-vector, $\mathbf{s}_{h\lambda}(x)'\equiv e_1'\left(\mathbf{A}_{n,h\lambda}(x)'\mathbf{A}_{n,h\lambda}(x)\right)^{-1}\mathbf{A}_{n,h\lambda}(x)'\mathbf{Z}_{n,h}(x)'\operatorname{diag}\left(\mathbf{k}_{h\lambda}(x)\right)$, $\mathbf{Z}_{n,h}(x) = \left(\mathbf{z}_{h,1}(x^c),\ldots,\mathbf{z}_{h,n}(x^c)\right)'$, and $\mathbf{k}_{h\lambda}(x)=\left(K_{h\lambda,1}(x),\ldots,K_{h\lambda,n}(x)\right)'$.

In the second step, we can estimate the parameter ρ_n^0 by the global IV method. Let $Z_n = (z_{n,1},\ldots,z_{n,n})'$. Let $\mathbf{S}_{h\lambda} = \left(\mathbf{s}_{h\lambda}\left(x_{n,1}\right),\ldots,\mathbf{s}_{h\lambda}\left(x_{n,n}\right)\right)'$, $\widetilde{Y}_n=\left(I_n-\mathbf{S}_{h\lambda}\right)Y_n$, and $\widetilde{\overline{Y}}_n=\left(I_n-\mathbf{S}_{h\lambda}\right)W_{1n}Y_n$. Then

$$\sum_{j=1}^{n}z_{n,j}\left\{y_{n,j}-\rho_n\bar{y}_{n,j}-\overline{m}_{\rho_n}\left(x_{n,j}\right)\right\}=Z_n'(\widetilde{Y}_n-\rho_n\widetilde{\overline{Y}}_n). \tag{2.14}$$

Let Ω_n be a symmetric $q\times q$ matrix that is positive semidefinite for large n. We can choose ρ_n to minimize

$$\left\|Z_n'(\widetilde{Y}_n-\rho_n\widetilde{\overline{Y}}_n)\right\|_{\Omega_n}, \tag{2.15}$$

where $\|A\|_{\Omega_n}\equiv\sqrt{A'\Omega_nA}$. It is easy to see the minimizer of (2.15) is given by

$$\widetilde{\rho}_n=\widetilde{\rho}_n\left(\Omega_n\right)=\frac{\widetilde{\overline{Y}}_n'Z_n\Omega_nZ_n'\widetilde{Y}_n}{\widetilde{\overline{Y}}_n'Z_n\Omega_nZ_n'\widetilde{\overline{Y}}_n}.$$

We will study the optimal choice of Ω_n for the given choice of $z_{n,j}$. (The optimal choice of $z_{n,j}$ is beyond the scope of this paper.) After we obtain $\widetilde{\rho}_n$, we can obtain the estimator of $\mathbf{M}_{\widetilde{h}}(x)$ by $\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x)\equiv\overline{\mathbf{M}}_{\widetilde{\rho}_n,\widetilde{h}\widetilde{\lambda}}(x)$ and that of $m(x)$ by $\widetilde{m}_{\widetilde{h}\widetilde{\lambda}}(x)\equiv\overline{m}_{\widetilde{\rho}_n,\widetilde{h}\widetilde{\lambda}}(x)=\mathbf{s}_{\widetilde{h}\widetilde{\lambda}}(x)'\left(Y_n-\widetilde{\rho}_nW_{1n}Y_n\right)$.

Note that we allow for different choices of smoothing parameters (h,λ) used in the estimation of ρ_n^0 and $\mathbf{M}_{\widetilde{h}}(x)$. We will explore the asymptotic properties of $\widetilde{\rho}_n$ and $\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x)$ in the next section.

3. Asymptotic theory

In this section, we study the asymptotic properties of $\widetilde{\rho}_n$ and $\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x)$.

3.1. Assumptions

To provide a rigorous asymptotic analysis, we maintain the following assumptions.

Assumption 1. (i) All diagonal elements $w_{1n,ii}$ of W_{1n} are zero. (ii) $\rho_n^0\in(-\underline{a}_{\rho n},\overline{a}_{\rho n})$ with $0<\underline{a}_{\rho n},\overline{a}_{\rho n}\le a_\rho<\infty$. (iii) The matrix $I_n-\rho W_{1n}$ is nonsingular for all $\rho\in(-\underline{a}_{\rho n},\overline{a}_{\rho n})$. (iv) The row and column sums of the sequences of matrices $\{W_{1n}\}$ and $\left\{\left(I_n-\rho_n^0W_{1n}\right)^{-1}\right\}$ are uniformly bounded in absolute value.

Assumption 1 concerns the essential features of spatial weights matrix. Assumption 1(i)–(iii) parallel Assumption 1(a)–(c) in Kelejian and Prucha (2010). Assumption 1(i) is clearly a normalization rule. Assumption 1(ii) concerns the parameter space of ρ_n^0 which may vary as the sample size changes. See Section 2.2 of Kelejian and Prucha (2010) for an excellent discussion on the parameter space for an autoregressive parameter. Assumption 1(iii) ensures that Y_n defined in (2.1) has the reduced form

$$Y_n=\left(I_n-\rho_n^0W_{1n}\right)^{-1}\mathbf{m}(X_n)+\left(I_n-\rho_n^0W_{1n}\right)^{-1}U_n. \tag{3.1}$$

Assumption 1(iv) parallels Assumption 5 of Lee (2004). Kelejian and Prucha (1998, 1999, 2001, 2010)also assume Assumption 1(iv) which limits the spatial correlation to some degree but facilitates the study of the asymptotic properties of the spatial parameter estimator.

Assumption 2. (i) $U_n=A_n\varepsilon_n$ such that the row and column sums of the sequences of matrices A_n are uniformly bounded in absolute value: $\sup_{1\le j\le n,n\ge1}\sum_{i=1}^{n}\left|a_{n,ij}\right|\le c_a$ and $\sup_{1\le i\le n,n\ge1}\sum_{j=1}^{n}\left|a_{n,ij}\right|\le c_a$ for some $c_a<\infty$. (ii) The error terms $\{\varepsilon_{n,i}:1\le i\le n,n\ge1\}$ satisfy: $E(\varepsilon_{n,i})=0$; $E(\varepsilon_{n,i}^2)=\sigma_{n,i}^2$ with $\sup_{1\le i\le n,n\ge1}\sigma_{n,i}^2\le\overline{\sigma}^2<\infty$; $\sup_{1\le i\le n,n\ge1}E\left|\varepsilon_{n,i}\right|^{4+\eta_1}\le\overline{\mu}_{4+\eta_1}<\infty$ for some small $\eta_1>0$. (iii) $\varepsilon_{n,1},\ldots,\varepsilon_{n,n}$ are totally independent.

Assumption 2 is fairly weak. It allows for not only heteroscedasticity but also spatial dependence in U_n. When $A_n=I_n$ such that $U_n=\varepsilon_n$, we have only heteroscedasticity in the error terms. In the presence of heteroscedasticity, the QMLE of Lee (2004) in the linear SAR models is generally inconsistent. For this reason, Kelejian and Prucha (2010) and Lin and Lee (2010) explore the GMM estimation of the linear SAR models with heteroscedasticity. They also require the existence of $(4+\eta_1)$th moments of $u_{n,i}$ or $\varepsilon_{n,i}$ for some $\eta_1>0$. More recently, Su and Yang (2009) study the instrumental variable quantile estimation of SAR models. They only require the existence of the first moment of $u_{n,i}$ but do not allow dependence among $u_{n,i}$, $i=1,2,\ldots,n$.

It is possible that U_n follows an SAR process, e.g., $U_n=\gamma_n^0W_{2n}U_n+\varepsilon_n$ with W_{2n} being a nonstochastic spatial weight matrix and γ_n^0 a spatial parameter in the error process. Under the condition that $\sup_n\left\|\gamma_n^0W_{2n}\right\|<1$, U_n has a reduced form: $U_n=A_n\varepsilon_n$ where $A_n\equiv(I_n-\gamma_n^0W_{2n})^{-1}$ will meet the conditions in Assumption 2(i). It is also possible that U_n forms an SMA process: $U_n=\gamma_n^0W_{2n}\varepsilon_n+\varepsilon_n$, in which case $A_n=I_n+\gamma_n^0W_{2n}$ will meet Assumption 2(i) if the row and column sums of W_{2n} are uniformly bounded in absolute value.

Assumption 3. (i) $x_{n,i}$, $i=1,\ldots,n$, are nonstochastic regressors with $x_{n,i}=(x_{n,i}^{c'},x_{n,i}^{d'})'\in\mathcal{X}_n^c\times\mathcal{X}_n^d\subset\mathbb{R}^p$, where $\mathcal{X}_n^c$ is a bounded set in $\mathbb{R}^{p_c}$ and $\mathcal{X}_n^d$ is the support of $x_{n,i}^d$ in $\mathbb{R}^{p_d}$ (the set of values that $\left\{x_{n,i}^d,i=1,\ldots,n,\right\}$ can take). (ii) There exist a function $\varphi_n\left(x^c,x^d\right)$ and a positive probability density/mass function $f_n\left(x^c,x^d\right)$ with support $\mathcal{X}_n\equiv\mathcal{X}_n^c\times\mathcal{X}_n^d$ such that

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}z_{n,i}^{(1)}\,v_n\left(x_{n,i}\right)\\ &=\lim_{n\to\infty}\sum_{x^d\in\mathcal{X}_n^d}\int_{\mathcal{X}_n^c}\varphi_n\left(x^c,x^d\right)v_n\left(x^c,x^d\right)f_n\left(x^c,x^d\right)dx^c\end{aligned} \tag{3.2}$$

for any bounded function $v_n\left(x^c,x^d\right)$ that is continuous in x^c, and $\varphi_n\left(x^c,x^d\right)$ and $f_n\left(x^c,x^d\right)$ are continuous in x^c and uniformly bounded on the support $\mathcal{X}_n$. (iii) $\lim_{n\to\infty}\varphi_n(x)=\varphi(x)$ and $\lim_{n\to\infty}f_n(x)=f(x)$ exist for each x on the support $\mathcal{X}\equiv\mathcal{X}^c\times\mathcal{X}^d$ of $f\left(x^c,x^d\right)$. (iv) $m\left(x^c,x^d\right)$ is second order continuously differentiable in x^c for each x^d on $\mathcal{X}_n^d$ and $\sup_{x\in\mathcal{X}}m(x)\le c_m<\infty$.

For notational simplicity, hereafter we will write $\int a_n(x)\,dx$ for $\sum_{x^d\in\mathcal{X}_n^d}\int_{\mathcal{X}_n^c} a_n\left(x^c, x^d\right) dx^c$ for any function a_n, where the summation is over all possible values of x^d on $\mathcal{X}_n^d$. The fixed bounded design assumption in Assumption 3(i) is typically assumed in the spatial econometrics literature, see Kelejian and Prucha (1998, 1999, 2001, 2010), Lee (2002, 2004) and Lin and Lee (2010), among others. Also, it allows for the fixed regressors to depend on n. Assumption 3(ii) and (iii) are typical in nonparametric regression with fixed regressor. For stationary random observations, $\varphi_n\left(x_{n,i}\right)$ can be regarded as the conditional expectation of $z_{n,i}^{(1)}$ given $x_{n,i}$. As in Linton (1995), Assumption 3(ii) does not preclude $\left\{x_{n,i}\right\}_{i=1}^n$ from being generated by some random mechanism. For example, if $x_{n,i}$ were i.i.d. with density $f_n(\cdot)$, then (3.2) holds with probability one. So even though we focus on the fixed regressor case, our analysis holds with probability one if $\left\{x_{n,i}\right\}_{i=1}^n$ are generated randomly, and in this case, we can interpret our analysis as being conditional on $\left\{x_{n,i}\right\}_{i=1}^n$. Assumption 3(iv) is required for the second order Taylor expansion of $m(x^c, x^d)$ with respect to x^c.

Assumption 4. (i) The kernel function $q(\cdot)$ is a continuous symmetric density function. There exists some small $\eta_2 > 0$ and a constant $c_q < \infty$ such that $\int |uq(u)|^{2+\eta_2} du < c_q$, $\int q^4(u)\,du < c_q$, $\sup_u q(u) < c_q$, and $\sup_u |u|\,q(u) < c_q$. (ii) As $n\to\infty$, $\|h\|\to 0$, $\|\lambda\|\to 0$, $\|\lambda\|$ is of the same order as $\|h\|^2$, $nh_1\cdots h_{p_c}\to\infty$, and $n\|h\|^4\to 0$. (iii) As $n\to\infty$, $\|\widetilde{h}\|\to 0$, $\|\widetilde{\lambda}\|\to 0$, $\|\widetilde{\lambda}\|$ is of the same order as $\|\widetilde{h}\|^2$, $n\widetilde{h}_1\cdots\widetilde{h}_{p_c}\to\infty$, and $n(\Pi_{s=1}^{p_c}\widetilde{h}_s)\|\widetilde{h}\|^4\to c_0\in[0,\infty)$.

Assumption 4(i) concerns the choice of kernel function. It is fairly standard in the nonparametric estimation literature. Assumption 4(ii) and (iii) concern the choice of smoothing parameters used in the first stage estimation of the parametric component (ρ_n^0) and the second stage estimation of the nonparametric component $(m(x))$, respectively. They are standard in the nonparametric regression with both continuous and discrete regressors with the only exception that undersmoothing is required for the bandwidth sequences used in the first stage. The conditions in Assumption 4(ii) imply implicitly that $p_c \le 3$. This is not too restrictive given the "curse of dimensionality" in the nonparametric literature. In the case where $p_c \ge 4$, one can apply higher order local polynomial estimation in place of the local linear procedure in the first stage.

Let $\Sigma_n \equiv \mathrm{Var}(\varepsilon_n) = \mathrm{diag}\left(\boldsymbol{\sigma}_n^2\right)$ with $\boldsymbol{\sigma}_n^2 = (\sigma_{n,1}^2,\ldots,\sigma_{n,n}^2)'$. Let $B_n \equiv n^{-1}Z_n'(I_n - \mathbf{S}_{h\lambda})W_{1n}Y_n$. The next assumption concerns the global instruments Z_n, and the weight matrix Ω_n.

Assumption 5. (i) $\Omega_n = \Omega + o_p(1)$ where Ω is positive semidefinite. (ii) The elements $z_{n,i}$ of Z_n are uniformly bounded such that $\sup_{1\le i\le n, n\ge 1}\left\|z_{n,i}\right\| \le c_z < \infty$ and $B_n = B + o_p(1)$ for a $q\times 1$ vector B with $B'\Omega B > 0$. (iii) $\Theta \equiv \lim_{n\to\infty} n^{-1}Z_n'(I_n - \mathbf{S}_{h\lambda}) A_n\Sigma_n A_n'\left(I_n - \mathbf{S}_{h\lambda}\right)' Z_n$ exists and $B'\Omega\Theta\Omega B > 0$.

Assumption 5(i) is standard and it allows the weight matrix Ω_n to be estimated from the data. The first part of Assumption 5(ii) is also standard in the spatial econometrics literature, whereas the second part of Assumption 5(ii) indicates the instruments relevance. Assumption 5(iii) allows Θ to be positive semidefinite which occurs if $z_{n,i}$ contains $x_{n,i}$.

3.2. Asymptotic property of $\widetilde{\rho}_n$

We first study the asymptotic property of $\widetilde{\rho}_n$. This is given in Theorem 3.1.

Theorem 3.1. *Under Assumptions 1–5,*

$$\sqrt{n}\left(\widetilde{\rho}_n - \rho_n^0\right) \xrightarrow{d} N\left(0, \left(B'\Omega B\right)^{-2} B'\Omega\Theta\Omega B\right).$$

The proof of the above theorem is tedious and is relegated to the appendix. Theorem 3.1 implies that the optimal choice of Ω is given by $\Omega = \Theta^+$ where Θ^+ is the Moore–Penrose generalized inverse of Θ. With this choice of weight, the asymptotic variance of $\widetilde{\rho}_n\left(\Theta^+\right)$ is minimized and given by $(B'\Theta^+B)^{-1}$. In the special case where $u_{n,i}$, $i = 1,\ldots,n$, are independent and homoscedastic such that $A_n = I_n$ and $\sigma_{n,i}^2 = \sigma_0^2$ for all i, it is easy to see that $\Theta = \sigma_0^2 \lim_{n\to\infty} n^{-1}Z_n'(I_n - \mathbf{S}_{h\lambda})(I_n - \mathbf{S}_{h\lambda})' Z_n$. For statistical inference, we need to estimate the asymptotic variance of $\sqrt{n}(\widetilde{\rho}_n - \rho_n^0)$. We postpone this study to Section 5.

3.3. Asymptotic property of $\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x)$

To consider the asymptotic property of $\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x)$, we define the following notation:

$$\mathbf{A}(x) \equiv \begin{pmatrix} \varphi(x) & \mathbf{0}_{q_1\times p_c} \\ \mathbf{0}_{q_1 p_c\times 1} & \kappa_{21}\varphi(x)\otimes I_{p_c}\end{pmatrix} \quad \text{and}$$

$$\Gamma_n(x) \equiv n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s \mathbf{Z}_{n,\widetilde{h}}'(x)\,\mathrm{diag}(\mathbf{k}_{\widetilde{h}\widetilde{\lambda}}(x))A_n\Sigma_n A_n' \times \mathrm{diag}(\mathbf{k}_{\widetilde{h}\widetilde{\lambda}}(x))\mathbf{Z}_{n,\widetilde{h}}(x),$$

where q_1 is the dimension of $z_{n,i}^{(1)}$, and $\kappa_{21} = \int s^2 q(s)\,ds$. Furthermore, for $s = 1,\ldots,p_d$, let

$$\mathbf{1}_s(x_{n,i}^d, x^d) = \mathbf{1}\left(x_{n,is}^d \ne x_s^d\right)\prod_{s'\ne s}^{p_d}\mathbf{1}\left(x_{n,is'}^d = x_{s'}^d\right)\mathbf{1}(s > p_1) + \mathbf{1}\left(|x_{n,is}^d - x_s^d| = 1\right)\prod_{s'\ne s}^{p_d}\mathbf{1}\left(x_{n,is'}^d = x_{s'}^d\right)\mathbf{1}(s\le p_1). \tag{3.3}$$

Define

$$b_{\widetilde{\lambda}}(x) = \sum_{v^d\in\mathcal{X}^d}\left[m\left(x^c, v^d\right) - m\left(x^c, x^d\right)\right] \times \sum_{s=1}^{p_d}\widetilde{\lambda}_s\mathbf{1}_s(v^d, x^d) f\left(x^c, v^d\right)\varphi\left(x^c, v^d\right). \tag{3.4}$$

That is, $\mathbf{1}_s(x_{n,i}^d, x^d) = 1$ requires that all the elements of $x_{n,i}^d$ but one should be identical to those of x^d. When $s > p_1$, $x_{n,is}^d$ is the unique element of $x_{n,i}^d$ that is not shared by x^d. When $s\le p_1$, $\mathbf{1}_s(x_{n,i}^d, x^d) = 1$ further requires that $|x_{n,is}^d - x_s^d| = 1$. As shown in the proof of Theorem 3.2, $b_{\widetilde{\lambda}}(x)$ will be a part of the asymptotic bias of our estimator of $m(x)$.

The following theorem shows that the nonparametric component $m(x)$ and its first derivatives with respect to its continuous component can be estimated at the regular nonparametric convergence rates.

Theorem 3.2. *Under Assumptions 1–5, if $\Gamma(x) \equiv \lim_{n\to\infty}\Gamma_n(x)$ exists and x^c is an interior point of $\mathcal{X}^c$, then*

$$\sqrt{n\prod_{s=1}^{p_c}\widetilde{h}_s}\left(\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x) - \mathbf{M}_{\widetilde{h}}(x) - \mathbf{A}^*(x) \times \begin{pmatrix}\frac{1}{2}\kappa_{21}\varphi(x)\sum_{s=1}^{p_c}\widetilde{h}_s^2 m_{ss}\left(x^c, x^d\right) + b_{\widetilde{\lambda}}(x) \\ \mathbf{0}_{q_1 p_c\times 1}\end{pmatrix}\right) \xrightarrow{d} N\left(0, \mathbf{A}^*(x)\,\Gamma(x)\,\mathbf{A}^*(x)'/f^2(x)\right),$$

where $\mathbf{M}_{\widetilde{h}}(x) = (m(x), (\widetilde{h} \odot \dot{m}(x))')'$, $m_{ss}(x^c, x^d) = \partial^2 m(x^c, x^d)/(x^c_{ss})^2$, $s = 1, \ldots, p_c$, $b_{\widetilde{\lambda}}(x)$ is defined in (3.4), and $\mathbf{A}^*(x) = (\mathbf{A}(x)' \mathbf{A}(x))^{-1} \mathbf{A}(x)'$.

Theorem 3.2 allows for both heteroscedasticity and spatial dependence in U_n. This general result does not come free. It requires that the limit of Γ_n should exist, which may not be easily verified for general specification of U_n. In the special case where $U_n = \varepsilon_n$ and $z^{(1)}_{n,i} = 1$, it is easy to see that

$$\Gamma = \lim_{n\to\infty} n^{-1} \prod_{s=1}^{p_c} \widetilde{h}_s \sum_{i=1}^{n} \times \begin{pmatrix} 1 & ((x^c_{n,i} - x^c)/\widetilde{h})' \\ (x^c_{n,i} - x^c)/\widetilde{h} & ((x^c_{n,i} - x^c)/\widetilde{h})((x^c_{n,i} - x^c)/\widetilde{h})' \end{pmatrix} \times K^2_{\widetilde{h}\widetilde{\lambda},i}(x)\, \sigma^2_{n,i}.$$

If further $\sigma^2_{n,i} = \sigma^2(x_{n,i})$, i.e., the error terms $u_{n,i}$ are conditionally heteroscedastic, it is easy to show that

$$\Gamma = f(x)\, \sigma^2(x) \begin{pmatrix} \kappa^{p_c}_{02} & 0 \\ 0 & \kappa^{p_c-1}_{02} \kappa_{22} I_{p_c} \end{pmatrix}, \tag{3.5}$$

where $\kappa_{ij} = \int s^i q^j(s)\, ds$ for $i, j = 0, 1, 2$. This implies the asymptotic independence between the estimator of $m(x)$ and that of its first derivative with respect to the continuous component of x.

4. Joint asymptotic distribution for the estimators of both spatial parameters in SARAR models

In this section we demonstrate that after we obtain the estimates of ρ^0_n and $m(x)$, we can also estimate the spatial parameter in the error term U_n. For clarity, we focus on the case where U_n also follows the SAR structure. In this case, we write

$$U_n = \gamma^0_n W_{2n} U_n + \varepsilon_n,$$

where W_{2n} is the spatial weight matrix in the error process that may be different from W_{1n}. Since the elements $u_{n,i}$ of U_n are not observed, we need to base our estimator of γ^0_n on a consistent estimator of U_n:

$$\widetilde{U}_n \equiv Y_n - \widetilde{\mathbf{m}}(X_n) - \widetilde{\rho}_n W_{1n} Y_n,$$

where $\widetilde{\mathbf{m}}(X_n) \equiv (\widetilde{m}(x_{n,1}), \ldots, \widetilde{m}(x_{n,n}))'$ and $\widetilde{m}(x) \equiv \widetilde{m}_{\widetilde{h}\widetilde{\lambda}}(x)$.

4.1. Estimation of the spatial parameter γ^0_n

In this subsection, we study the consistent estimation of the spatial parameter γ_n. Following the literature, we assume that the diagonal elements of W_{2n} are zero. Let $\overline{\varepsilon}_n = W_{2n} \varepsilon_n$. Then we have

$$n^{-1} E\left[\overline{\varepsilon}'_n \overline{\varepsilon}_n\right] = n^{-1} \mathrm{tr}\left\{W_{2n} \Sigma_n W'_{2n}\right\} \quad \text{and} \quad n^{-1} E\left[\overline{\varepsilon}'_n \varepsilon_n\right] = 0. \tag{4.1}$$

Like Kelejian and Prucha (2010), it is convenient to rewrite the above moment conditions as

$$n^{-1} E\begin{bmatrix} \varepsilon'_n A_{1n} \varepsilon_n \\ \varepsilon'_n A_{2n} \varepsilon_n \end{bmatrix} = 0, \tag{4.2}$$

where $A_{1n} \equiv W'_{2n} W_{2n} - \mathrm{diag}(\mathbf{w}_{2n})$, $A_{2n} \equiv (W_{2n} + W'_{2n})/2$, and $\mathbf{w}_{2n} \equiv (w'_{2n,\cdot 1} w_{2n,\cdot 1}, \ldots, w'_{2n,\cdot n} w_{2n,\cdot n})'$ with $w_{2n,\cdot i}$ denoting the ith column of W_{2n}. We will estimate γ^0_n based upon the moment conditions in (4.2).

Noting that $\varepsilon_n = \left(I_n - \gamma^0_n W_{2n}\right) U_n$, we can substitute this expression into (4.2) to yield

$$\psi_n - \Psi_n \theta^0_n = 0, \tag{4.3}$$

where $\theta^0_n \equiv [\gamma^0_n, (\gamma^0_n)^2]'$,

$$\psi_n \equiv \begin{bmatrix} \psi_{n,1} \\ \psi_{n,2} \end{bmatrix} = \begin{bmatrix} n^{-1} E\left(U'_n A_{1n} U_n\right) \\ n^{-1} E\left(U'_n A_{2n} U_n\right) \end{bmatrix}, \quad \text{and} \tag{4.4}$$

$$\Psi_n \equiv \begin{bmatrix} \psi_{n,11} & \psi_{n,12} \\ \psi_{n,21} & \psi_{n,22} \end{bmatrix} = \begin{bmatrix} 2n^{-1} E\left(U'_n W'_{2n} A_{1n} U_n\right) & -n^{-1} E\left(U'_n W'_{2n} A_{1n} W_{2n} U_n\right) \\ 2n^{-1} E\left(U'_n W'_{2n} A_{2n} U_n\right) & -n^{-1} E\left(U'_n W'_{2n} A_{2n} W_{2n} U_n\right) \end{bmatrix}. \tag{4.5}$$

We then obtain the estimators $\widetilde{\Psi}_n = \left[\widetilde{\psi}_{n,ij}\right]_{i,j=1,2}$ and $\widetilde{\psi}_n = [\widetilde{\psi}_{n,1}, \widetilde{\psi}_{n,2}]'$ for the elements of $\Psi_n = \left[\psi_{n,ij}\right]_{i,j=1,2}$ and $\psi_n = [\psi_{n,1}, \psi_{n,2}]'$ by suppressing the expectation operator and replacing the disturbance vector U_n by $\widetilde{U}_n$. Let $q_n(\gamma_n) = \widetilde{\psi}_n - \widetilde{\Psi}_n \theta_n$, where $\theta_n \equiv [\gamma_n, \gamma^2_n]'$. We obtain the generalized methods of moments estimator $\widetilde{\gamma}_n \equiv \widetilde{\gamma}_n(\Upsilon_n)$ for γ^0_n by minimizing the following objective function

$$Q_n \equiv q_n(\gamma_n)' \Upsilon_n q_n(\gamma_n), \tag{4.6}$$

where Υ_n is a 2×2 symmetric positive semidefinite matrix.

As shown in the appendix (see (C.5) and (C.7)), the dominant term of $\sqrt{n}\left(\widetilde{\gamma}_n - \gamma^0_n\right)$ can be written as a linear combination of

$$v_n \equiv n^{-1/2} \begin{bmatrix} \varepsilon'_n A_{1n} \varepsilon_n + a'_{1n} \varepsilon_n \\ \varepsilon'_n A_{2n} \varepsilon_n + a'_{2n} \varepsilon_n \end{bmatrix}, \tag{4.7}$$

where for $k = 1, 2$,

$$a'_{kn} = -n^{-1} E[\varepsilon'_n \overline{C}_{kn} \varepsilon_n] \times (B'\Omega B)^{-1} B'\Omega Z'_n \left(I_n - \mathbf{S}_{h\gamma}\right) \left(I_n - \gamma^0_n W_{2n}\right)^{-1}, \tag{4.8}$$

$$G_{1n} \equiv W_{1n} \left(I_n - \rho^0_n W_{1n}\right)^{-1},$$

$$\overline{C}_{kn} = 2\left(I_n - \gamma^0_n W'_{2n}\right)^{-1} G'_{1n} \left(I_n - \gamma^0_n W'_{2n}\right) A_{kn}. \tag{4.9}$$

Noticing that the diagonal elements of the matrices $A_{kn}(k = 1, 2)$ are zero, we can apply Theorem A.1 to deduce that the asymptotic variance-covariance (VC) matrix of the vector of linear quadratic forms in (4.7) is given by $\Phi_{n,\gamma\gamma} = [\phi_{n\gamma\gamma,kl}]_{k,l=1,2}$, where

$$\phi_{n\gamma\gamma,kl} = 2n^{-1} \mathrm{tr}\left[A_{kn} \Sigma_n A_{ln} \Sigma_n\right] + n^{-1} a'_{kn} \Sigma_n a_{ln}. \tag{4.10}$$

To state the next theorem, we add the following assumption.

Assumption 6. (i) All diagonal elements $w_{2n,ii}$ of W_{2n} are zero. (ii) $\gamma^0_n \in (-\underline{a}_{\gamma n}, \overline{a}_{\gamma n})$ with $0 < \underline{a}_{\gamma n}, \overline{a}_{\gamma n} \le a_\gamma < \infty$. (iii) The matrix $I_n - \gamma W_{2n}$ is nonsingular for all $\gamma \in (-\underline{a}_{\gamma n}, \overline{a}_{\gamma n})$. (iv) The row and column sums of the sequences of matrices $\{W_{2n}\}$ and $\{\left(I_n - \gamma^0_n W_{2n}\right)^{-1}\}$ are uniformly bounded in absolute value.

Assumption 6 parallels Assumption 1 so that a discussion similar to that after Assumption 1 also applies here. The following theorem concerns the asymptotic normal distribution of $\widetilde{\gamma}_n$.

Theorem 4.1. *Suppose that Assumptions 1–6 hold. Furthermore, suppose that $n\|\widetilde{h}\|^8 \to 0$ and $n(\prod_{s=1}^{p_c} \widetilde{h}_s \log n)^2 \to \infty$ as $n \to \infty$. Suppose that $\lambda_{\min}\left(\Psi'_n \Psi_n\right) \ge c_\psi > 0$, $\lambda_{\min}(\Upsilon_n) \ge c_\Upsilon > 0$ and $\lambda_{\min}\left(\Phi_{n,\gamma\gamma}\right) \ge c_{\Phi_{\gamma\gamma}} > 0$. Then*

$$\sqrt{n}\left(\widetilde{\gamma}_n - \gamma^0_n\right) = \left(J'_n \Upsilon_n J_n\right)^{-1} J'_n \Upsilon_n \Phi^{1/2}_{n,\gamma\gamma} \xi_n + o_p(1) \xrightarrow{d} N\left(0, \Omega_\gamma\right),$$

where $J_n = \Psi_n[1, 2\gamma^0_n]'$, $\Omega_\gamma = \lim_{n\to\infty} \left(J'_n \Upsilon_n J_n\right)^{-1} J'_n \Upsilon_n \Phi_{n,\gamma\gamma} \Upsilon_n J_n \left(J'_n \Upsilon_n J_n\right)^{-1} \ge c > 0$, and $\xi_n \equiv \Phi^{-1/2}_{n,\gamma\gamma} v_n \xrightarrow{d} N(0, I_2)$.

Like Assumption 4(ii), the extra conditions on $\widetilde{h}$ in Theorem 4.1 require that $p_c \leq 3$. As p_c increases, it becomes more and more difficult to estimate $m(x)$ so that the distance between $\widetilde{m}(x_{n,i})$ and $m(x_{n,i})$ gets larger and larger. Unless we can estimate $m(x)$ at a sufficiently accurate rate, we cannot estimate the spatial parameter γ_n^0 at the usual $\sqrt{n}$-parametric rate. From the proof of Lemma C.1 in the appendix, we need the bias of $\widetilde{m}(x_{n,i})$ to be $o(n^{-1/4})$, which is a typical requirement in the semiparametric literature where nonparametric estimation is conducted before one obtains a parametric estimator.

Clearly, the choice $\Upsilon_n = \Phi_{n,\gamma\gamma}^{-1}$ minimizes Ω_γ, and in this case, $\Omega_\gamma = \lim_{n\to\infty}\left(J_n'\Phi_{n,\gamma\gamma}^{-1}J_n\right)^{-1}$. In practice, Υ_n may not be observable and one can replace Υ_n by its consistent estimate in (4.6) without altering the asymptotic results in Theorem 4.1.

4.2. Joint asymptotic distribution of $\widetilde{\rho}_n$ and $\widetilde{\gamma}_n$

In this subsection, we study the joint asymptotic distribution of the estimators for ρ_n^0 and γ_n^0. Let $\delta_n^0 \equiv (\rho_n^0, \gamma_n^0)'$ and $\widetilde{\delta}_n \equiv (\widetilde{\rho}_n, \widetilde{\gamma}_n)'$. In light of the proofs of Theorems 3.1 and 4.1,

$$\sqrt{n}\left(\widetilde{\rho}_n - \rho_n^0\right) = \left(B'\Omega B\right)^{-1} B'\Omega n^{-1/2} P_n'\varepsilon_n + o_p(1), \quad \text{and}$$
$$\sqrt{n}\left(\widetilde{\gamma}_n - \gamma_n^0\right) = \left(J_n'\Upsilon_n J_n\right)^{-1} J_n'\Upsilon_n v_n + o_p(1),$$

where $P_n' = Z_n'(I_n - \mathbf{S}_{h\lambda})\left(I_n - \gamma_n^0 W_{2n}\right)^{-1}$, and v_n is defined in (4.7). Hence, the joint limiting distribution of $\sqrt{n}\left(\widetilde{\rho}_\Omega - \rho_n^0\right)$ and $\sqrt{n}\left(\widetilde{\gamma}_n - \gamma_n^0\right)$ will depend on the limiting distribution of $v_n^a \equiv [n^{-1/2}(P_n'\varepsilon_n)', v_n']'$. Observe that v_n^a is a vector of linear and linear quadratic forms studied in Appendix A. Its variance-covariance matrix is given by

$$\Phi_n \equiv \begin{bmatrix} \Phi_{n,\rho\rho} & \Phi_{n,\rho\gamma} \\ \Phi_{n,\rho\gamma}' & \Phi_{n,\gamma\gamma} \end{bmatrix}, \tag{4.11}$$

where $\Phi_{n,\rho\rho} = n^{-1}P_n'\Sigma_n P_n$, $\Phi_{n,\rho\gamma} = n^{-1}P_n'\Sigma_n[a_{1n}, a_{2n}]$, and elements of $\Phi_{n,\gamma\gamma}$ are defined in (4.10).

The following theorem establishes the joint asymptotic normality of $\widetilde{\rho}_n$ and $\widetilde{\gamma}_n$.

Theorem 4.2. *Suppose that the conditions in Theorem 4.1 hold. Suppose that $\lambda_{\min}(\Phi_n) \geq c_\Phi > 0$. Then*

$$\sqrt{n}\left(\widetilde{\delta}_n - \delta_n^0\right) = \begin{bmatrix} \left(B'\Omega B\right)^{-1} B'\Omega & 0 \\ 0 & \left(J_n'\Upsilon_n J_n\right)^{-1} J_n'\Upsilon_n \end{bmatrix} \times \Phi_n^{1/2}\xi_n^a + o_p(1) \xrightarrow{d} N(0, \Omega_\delta),$$

where $\Omega_\delta \equiv \lim_{n\to\infty}\Omega_{n,\delta}$, $\Omega_{n,\delta} \equiv H_n'\Phi_n H_n$, $H_n \equiv \begin{bmatrix} \Omega B\left(B'\Omega B\right)^{-1} & 0 \\ 0 & \Upsilon_n J_n\left(J_n'\Upsilon_n J_n\right)^{-1} \end{bmatrix}$, and $\xi_n^a \equiv \Phi_n^{-1/2}v_n^a \xrightarrow{d} N(0, I_{q+2})$.

Theorem 4.2 implies that we can make joint statistical inference on both spatial parameters ρ_n^0 and γ_n^0. To do so, consistent estimation of the asymptotic VC matrix Ω_δ is needed. See the next section for the exploration.

4.3. An alternative method: joint estimation of ρ_n^0 and γ_n^0

In this subsection, we study the joint estimation of the spatial parameters ρ_n^0 and γ_n^0 based on the moment conditions in (4.2).[2]

Let $\varepsilon_n(\delta_n) \equiv (I_n - \gamma_n W_{2n})(Y_n - \widetilde{\mathbf{m}}(X_n) - \rho_n W_{1n}Y_n)$ where $\delta_n \equiv (\rho_n, \gamma_n)'$ and $\widetilde{\mathbf{m}}(X_n) \equiv \left(\widetilde{m}_{\widetilde{h}\widetilde{\lambda}}(x_{n,1}), \ldots, \widetilde{m}_{\widetilde{h}\widetilde{\lambda}}(x_{n,n})\right)'$. Let $q_n^*(\delta_n) \equiv [q_{1n}^*(\delta_n), q_{2n}^*(\delta_n)]'$, where $q_{kn}^*(\delta_n) \equiv n^{-1}\varepsilon_n(\delta_n)' A_{kn}\varepsilon_n(\delta_n)$ for $k = 1, 2$. We propose to estimate δ_n^0 by minimizing the following objective function

$$Q_n^* \equiv q_n^*(\delta_n)' \Upsilon_n^* q_n^*(\delta_n), \tag{4.12}$$

where Υ_n^* is a 2×2 matrix that is positive definite for sufficiently large n. Let $\widehat{\delta}_n \equiv (\widehat{\rho}_n, \widehat{\gamma}_n)'$ denote the solution to the above minimization problem. As shown in the appendix, $\partial q_n^*\left(\widehat{\delta}_n\right)/\partial\delta_n' = J_n^* + o_p(1)$, where J_n^* is given in Box I. The dominant term of $\sqrt{n}\left(\widehat{\delta}_n - \delta_n^0\right)$ can be written as a linear combination of $v_n^* \equiv n^{-1/2}[\varepsilon_n'A_{1n}\varepsilon_n\ \varepsilon_n'A_{2n}\varepsilon_n]'$. Then Theorem A.1 implies that the asymptotic VC matrix of v_n^* is given by $\Phi_n^* \equiv [\phi_{n,kl}^*]_{k,l=1,2}$, where $\phi_{n,kl}^* \equiv 2n^{-1}\mathrm{tr}(A_{kn}\Sigma_n A_{ln}\Sigma_n)$.

Let $\overline{q}_n^*(\delta_n) \equiv [\overline{q}_{1n}^*(\delta_n), \overline{q}_{2n}^*(\delta_n)]'$, where $\overline{q}_{kn}^*(\delta_n) \equiv n^{-1}E[\overline{\varepsilon}_n(\delta_n)' A_{kn}\overline{\varepsilon}_n(\delta_n)]$ and $\overline{\varepsilon}_n(\delta_n) \equiv (I_n - \gamma_n W_{2n})(Y_n - \mathbf{m}(X_n) - \rho_n W_{1n}Y_n)$. The following theorem concerns the asymptotic normal distribution of $\widehat{\delta}_n$.

Theorem 4.3. *Suppose that Assumptions 1–6 hold. Suppose that $n\|\widetilde{h}\|^8 \to 0$ and $n(\prod_{s=1}^{p_c}\widetilde{h}_s \log n)^2 \to \infty$ as $n \to \infty$. Suppose that $\inf_{\{\delta_n:\|\delta_n-\delta_n^0\|>\epsilon\}}\left\|\overline{q}_n^*(\delta_n)\right\| \geq c_{q^*} > 0$ for each $\epsilon > 0$, $\lambda_{\min}\left(J_n^{*\prime}J_n^*\right) \geq c_{J^*} > 0$, $\lambda_{\min}\left(\Upsilon_n^*\right) \geq c_{\Upsilon^*} > 0$ and $\lambda_{\min}\left(\Phi_n^*\right) \geq c_{\Phi^*} > 0$. Then*

$$\sqrt{n}\left(\widehat{\delta}_n - \delta_n^0\right) = \left(J_n^{*\prime}\Upsilon_n^* J_n^*\right)^{-1} J_n^{*\prime}\Upsilon_n^*\Phi_n^{*1/2}\xi_n^* + o_p(1) \xrightarrow{d} N\left(0, \Omega_\delta^*\right),$$

where $\Omega_\delta^* = \lim_{n\to\infty}\left(J_n^{*\prime}\Upsilon_n^* J_n^*\right)^{-1} J_n^{*\prime}\Upsilon_n^*\Phi_n^*\Upsilon_n^* J_n^*\left(J_n^{*\prime}\Upsilon_n^* J_n^*\right)^{-1}$, and $\xi_n \equiv \Phi_n^{*-1/2}v_n^* \xrightarrow{d} N(0, I_2)$.

Similar remarks to those after Theorem 4.1 hold. In particular, the choice $\Upsilon_n^* = \Phi_n^{*-1}$ minimizes Ω_δ^* asymptotically, and in this case $\Omega_\delta^* = \lim_{n\to\infty}\left(J_n^{*\prime}\Phi_n^{*-1}J_n^*\right)^{-1}$.

Even though it is hard to compare the asymptotic VC matrix of $\widehat{\delta}_n$ with that of $\widetilde{\delta}_n$, a noticeable difference is that the former does not depend on matrices like B and Ω. This implies that the preliminary estimation of the nonparametric component plays asymptotically negligible role in the joint estimation of ρ_n^0 and γ_n^0. In addition, it is easy to verify that the condition on J_n^* is violated when $\rho_n^0 = \gamma_n^0 = 0$ and $W_{1n} = W_{2n}$. In this case, the asymptotic VC matrix of $\widehat{\delta}_n$ is singular. For this reason, we only implement $\widetilde{\delta}_n$ in Section 6.

5. Estimation of the asymptotic VC matrices

In this section, we discuss the estimation of the asymptotic variance-covariance matrices for both estimators of the parametric and nonparametric components in the model. In particular, we focus on the consistent estimation of the two asymptotic covariance matrices in Theorems 3.2 and 4.2.

We first define a HAC type estimator for Φ_n defined in (4.11). For this purpose, let $\widetilde{\varepsilon}_n \equiv (I_n - \widetilde{\gamma}_n W_{2n})\widetilde{U}_n = (\widetilde{\varepsilon}_{n,1}, \ldots, \widetilde{\varepsilon}_{n,n})'$, $\widetilde{\boldsymbol{\varepsilon}}_n^2 \equiv (\widetilde{\varepsilon}_{n,1}^2, \ldots, \widetilde{\varepsilon}_{n,n}^2)'$, $\widetilde{\Sigma}_n \equiv \mathrm{diag}(\widetilde{\boldsymbol{\varepsilon}}_n^2)$, $\widetilde{a}_{kn}' \equiv -n^{-1}\widetilde{\varepsilon}_n'\widetilde{C}_{kn}\widetilde{\varepsilon}_n(B_n'\Omega_n B_n)^{-1} B_n'\Omega_n Z_n'\left(I_n - \mathbf{S}_{h\gamma}\right)(I_n - \widetilde{\gamma}_n W_{2n})^{-1}$, and $\widetilde{C}_{kn} \equiv 2\left(I_n - \widetilde{\gamma}_n W_{2n}'\right)^{-1}\left(I_n - \widetilde{\rho}_n W_{1n}'\right)^{-1} W_{1n}'\left(I_n - \widetilde{\gamma}_n W_{2n}'\right) A_{kn}$ where $k = 1, 2$. We also need to specify an estimator for P_n : $\widetilde{P}_n \equiv Z_n'(I_n - \mathbf{S}_{h\lambda}) \times (I_n - \widetilde{\gamma}_n W_{2n})^{-1}$. Let

$$\widetilde{\Phi}_n \equiv \begin{bmatrix} \widetilde{\Phi}_{n,\rho\rho} & \widetilde{\Phi}_{n,\rho\gamma} \\ \widetilde{\Phi}_{n,\rho\gamma}' & \widetilde{\Phi}_{n,\gamma\gamma} \end{bmatrix},$$

[2] More generally, one could consider the GMM estimation of ρ_n^0 and γ_n^0 based on both linear and quadratic moment conditions for ε_n as in Liu et al. (2010).

$$J_n^* = 2n^{-1}\begin{bmatrix} E\left\{U_n' G_{1n}'\left(I_n-\gamma_n^0 W_{2n}\right)' A_{1n}\left(I_n-\gamma_n^0 W_{2n}\right) U_n\right\} & E\left\{U_n' W_{2n}' A_{1n}\left(I_n-\gamma_n^0 W_{2n}\right) U_n\right\} \\ E\left\{U_n' G_{1n}'\left(I_n-\gamma_n^0 W_{2n}\right)' A_{2n}\left(I_n-\gamma_n^0 W_{2n}\right) U_n\right\} & E\left\{U_n' W_{2n}' A_{2n}\left(I_n-\gamma_n^0 W_{2n}\right) U_n\right\}\end{bmatrix} \tag{4.13}$$

Box I.

$$\widetilde{J}_n^* \equiv 2n^{-1}\begin{bmatrix} \widehat{U}_n'\widehat{G}_{1n}'\left(I_n-\widehat{\gamma}_n W_{2n}\right)' A_{1n}\left(I_n-\widehat{\gamma}_n W_{2n}\right)\widehat{U}_n & \widehat{U}_n' W_{2n}' A_{1n}\left(I_n-\widehat{\gamma}_n W_{2n}\right)\widehat{U}_n \\ \widehat{U}_n'\widehat{G}_{1n}'\left(I_n-\widehat{\gamma}_n W_{2n}\right)' A_{2n}\left(I_n-\widehat{\gamma}_n W_{2n}\right)\widehat{U}_n & \widehat{U}_n' W_{2n}' A_{2n}\left(I_n-\widehat{\gamma}_n W_{2n}\right)\widehat{U}_n\end{bmatrix}$$

Box II.

with $\widetilde{\Phi}_{n,\rho\rho} = n^{-1}\widetilde{P}_n'\widetilde{\Sigma}_n\widetilde{P}_n$, $\widetilde{\Phi}_{n,\rho\gamma} = n^{-1}\widetilde{P}_n'\widetilde{\Sigma}_n[\widetilde{a}_{1n}, \widetilde{a}_{2n}]$, $\widetilde{\Phi}_{n,\gamma\gamma} = (\widetilde{\phi}_{n\gamma\gamma,kl})_{k,l=1,2}$, where $\widetilde{\phi}_{n\gamma\gamma,kl} = 2n^{-1}\mathrm{tr}(A_{kn}\widetilde{\Sigma}_n A_{ln}\widetilde{\Sigma}_n) + n^{-1}\widetilde{a}_{kn}'\widetilde{\Sigma}_n\widetilde{a}_{ln}$. Let $\widetilde{J}_n \equiv \widetilde{\Psi}_n[1, 2\widetilde{\gamma}_n]$ and $\widetilde{\Omega}_{n,\delta} \equiv \widetilde{H}_n'\widetilde{\Phi}_n\widetilde{H}_n$, where

$$\widetilde{H}_n \equiv \begin{bmatrix} \Omega_n B_n\left(B_n'\Omega_n B_n\right)^{-1} & 0 \\ 0 & \Upsilon_n\widetilde{J}_n(\widetilde{J}_n'\Upsilon_n\widetilde{J}_n)^{-1}\end{bmatrix}.$$

The following theorem says that we can consistently estimate Ω_δ by $\widetilde{\Omega}_{n,\delta}$.

Theorem 5.1. *Suppose the conditions in Theorem 4.2 hold. Suppose in addition that* $\sup_n\left|\rho_n^0\right| < \rho_*$ *and* $\sup_n\left|\gamma_n^0\right| < \gamma_*$ *such that* $\|\rho_* W_{1n}\|_\infty < 1$ *and* $\|\gamma_* W_{2n}\|_\infty < 1$, *where* $\|\cdot\|_\infty$ *denotes the maximum row sum norm. Then* $\widetilde{\Phi}_n - \Phi_n = o_p(1)$ *and* $\widetilde{\Omega}_{n,\delta} - \Omega_{n,\delta} = o_p(1)$.

Let $\widehat{U}_n \equiv Y_n - \widetilde{\mathbf{m}}(X_n) - \widehat{\rho}_n W_{1n}Y_n$, $\widehat{\varepsilon}_n \equiv (I_n - \widehat{\gamma}_n W_{2n})\widehat{U}_n = (\widehat{\varepsilon}_{n,1}, \ldots, \widehat{\varepsilon}_{n,n})'$, $\widehat{\boldsymbol{\varepsilon}}_n^2 \equiv (\widehat{\varepsilon}_{n,1}^2, \ldots, \widehat{\varepsilon}_{n,n}^2)'$, and $\widehat{\Sigma}_n \equiv \mathrm{diag}(\widehat{\boldsymbol{\varepsilon}}_n^2)$. Let $\widehat{G}_{1n} \equiv W_{1n}(I_n - \widehat{\rho}_n W_{1n})^{-1}$ and $\widetilde{\Phi}_n^* \equiv [\widetilde{\phi}_{n,kl}^*]_{k,l=1,2}$ where $\widetilde{\phi}_{n,kl}^* \equiv 2n^{-1}\mathrm{tr}(A_{kn}\widehat{\Sigma}_n A_{ln}\widehat{\Sigma}_n)$. Define $\widetilde{J}_n^*$ in Box II. Let $\widetilde{\Omega}_{n,\delta}^* \equiv \left(\widetilde{J}_n^{*\prime}\Upsilon_n^*\widetilde{J}_n^*\right)^{-1}\widetilde{J}_n^*\Upsilon_n^*\widetilde{\Phi}_n^*\Upsilon_n^*\widetilde{J}_n^*\left(\widetilde{J}_n^{*\prime}\Upsilon_n^*\widetilde{J}_n^*\right)^{-1}$. The following theorem establishes the consistency of $\widetilde{\Omega}_{n,\delta}^*$ for Ω_δ^*.

Theorem 5.2. *Suppose the conditions in Theorem 4.3 hold. Suppose in addition that* $\sup_n\left|\rho_n^0\right| < \rho_*$ *and* $\sup_n\left|\gamma_n^0\right| < \gamma_*$ *such that* $\|\rho_* W_{1n}\|_\infty < 1$ *and* $\|\gamma_* W_{2n}\|_\infty < 1$. *Then* $\widetilde{\Phi}_n^* - \Phi_n^* = o_p(1)$ *and* $\widetilde{\Omega}_{n,\delta}^* - \Omega_{n,\delta}^* = o_p(1)$.

Statistical inference associated with $\delta_n^0 \equiv (\rho_n^0, \gamma_n^0)'$ can now be conducted based upon the above theorems and Theorems 4.2 and 4.3. For example, we can test the hypothesis of zero spatial correlation in both the regression model and the disturbance term, i.e., $H_0: \rho_n^0 = \gamma_n^0 = 0$. As special cases, we can also test the simple null hypothesis $H_0: \rho_n^0 = 0$ or $H_0: \gamma_n^0 = 0$.

In practice, we may be also interested in making a statistical inference on the nonparametric component. For this purpose, we need to estimate the asymptotic VC matrix in Theorem 3.2. Let $\widetilde{\Gamma}_n(x) \equiv n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s\mathbf{Z}_{n,\widetilde{h}}'(x)\,\mathrm{diag}(\mathbf{k}_{\widetilde{h}\widetilde{\lambda}}(x))A_n\widetilde{\Sigma}_n A_n'\mathrm{diag}(\mathbf{k}_{\widetilde{h}\widetilde{\lambda}}(x))\,\mathbf{Z}_{n,\widetilde{h}}(x)$, $\widetilde{f}_n(x) \equiv n^{-1}\sum_{i=1}^n K_{\widetilde{h}\widetilde{\lambda},i}(x)$, and $\widetilde{\varphi}_n(x) \equiv n^{-1}\sum_{i=1}^n z_{n,i}^{(1)}K_{\widetilde{h}\widetilde{\lambda},i}(x)/\widetilde{f}_n(x)$. Define $\widetilde{\mathbf{A}}_n(x)$ as $\mathbf{A}(x)$ with $\widetilde{\varphi}_n(x)$ replacing $\varphi_n(x)$. Let $\widetilde{\mathbf{A}}_n^*(x) = (\widetilde{\mathbf{A}}_n(x)'\widetilde{\mathbf{A}}_n(x))^{-1}\widetilde{\mathbf{A}}_n(x)'$, and $\widetilde{\Omega}_{n,M}(x) \equiv \widetilde{\mathbf{A}}_n^*(x)\widetilde{\Gamma}_n(x)\widetilde{\mathbf{A}}_n^*(x)'/\widetilde{f}_n(x)^2$. The following result is sufficient to establish the consistency of the estimator for the asymptotic VC matrix.

Theorem 5.3. *Suppose that the conditions in Theorem 3.2 hold. Then* $\widetilde{\Gamma}_n(x) - \Gamma_n(x) = o_p(1)$, $\widetilde{f}_n(x) - f_n(x) = o(1)$, $\widetilde{\varphi}_n(x) - \varphi_n(x) = o(1)$, *and* $\widetilde{\Omega}_{n,M}(x) - \Omega_{n,M}(x) = o_p(1)$, *where* $\Omega_{n,M}(x) \equiv \mathbf{A}^*(x)\,\Gamma(x)\,\mathbf{A}^*(x)/f_n^2(x)$.

In proving the above theorem, we assume that A_n is known. In the case of spatial autoregression error, $A_n = I_n - \gamma_n^0 W_{2n}$, which is not observed. It is easy to show that the replacement of A_n by, say, $\widetilde{A}_n \equiv I_n - \widetilde{\gamma}_n W_{2n}$ in the definition of $\widetilde{\Gamma}_n$ will have asymptotically negligible effect on the consistent result in Theorem 5.3 provided Assumption 6 is also satisfied.

6. Monte Carlo simulations

We now present a small set of Monte Carlo experiments to examine the finite sample performance of our semiparametric GMM estimators. Like Su and Yang (2009), we generate the spatial weight matrix $W_n \equiv W_{1n}$ according to Rook contiguity, by randomly allocating the n spatial units on a lattice of $5 \times m$ ($\geq n$) squares, finding the neighbors for each unit, and then row normalizing.

We generate the data from the following data generating processes (DGPs):

DGP 1: $Y_n = 1 + X_{n1}^d + X_{n2}^d + X_{n1}^c + \rho_n^0 W_n Y_n + X_{n1}^c \odot X_{n1}^c + U_n$,

DGP 2: $Y_n = 1 + X_{n1}^d + X_{n2}^d + X_{n1}^c + \rho_n^0 W_n Y_n + 0.5X_{n1}^d \odot \exp(X_{n1}^c) + 0.5X_{n2}^d \odot X_{n1}^c \odot X_{n1}^c + U_n$,

DGP 3: $Y_n = 1 + X_{n1}^d + X_{n2}^d + X_{n1}^c + X_{n2}^c + \rho_n^0 W_n Y_n + X_{n1}^d \odot \cos\left(0.5\pi X_{n1}^c\right) + X_{n2}^d \odot \sin(0.5\pi X_{n2}^c) + 0.5X_{n1}^c \odot X_{n2}^c + U_n$,

where $\odot$ denotes the Hadamard product, $X_{n1}^c = \left(x_{n,11}^c, \ldots, x_{n,1n}^c\right)'$, $X_{n2}^c = \left(x_{n,21}^c, \ldots, x_{n,2n}^c\right)'$, $x_{n,1i}^c$'s are i.i.d. and each is equal to the sum of 48 independent random variables each uniformly distributed on $[-0.25, 0.25]$, and $x_{n,2i}^c$'s are i.i.d. $U(-2, 2)$; for $t = 1, 2$, $X_{nt}^d = \left(x_{n,t1}^d, \ldots, x_{n,tn}^d\right)'$, $P(x_{n,t1}^d = l) = 0.5$ for $l = 0, 1$. According to the central limit theorem, we can treat $x_{n,1i}^c$ as being nearly a normal random variable with truncated support on $[-12, 12]$. The error term is generated according to the SAR process: $U_n = \gamma_n^0 W_n U_n + \varepsilon_n$, where $\varepsilon_n = \left(\varepsilon_{n,1}, \ldots, \varepsilon_{n,n}\right)'$, $\varepsilon_{n,i} = \sqrt{0.5(1 + x_{n,1i}^{c2})}\eta_i$, and η_i's are i.i.d. $N(0, 1)$. We will consider different population values of $\left(\rho_n^0, \gamma_n^0\right)$: $(0, 0)$, $(0.3, 0)$, $(0, 0.3)$, and $(0.3, 0.3)$.

To implement our estimation procedure, we need to choose the kernel function and bandwidth sequences. Throughout, we will choose a Gaussian kernel (for DGPs 1 and 2) or the product of a Gaussian kernel (for DGP 3). That is, for DGPs 1 and 2 where there is only one continuous exogenous regressor, we choose $q(x) = (2\pi)^{-1/2}\exp\left(-x^2/2\right)$; for DGP 3 where there are two non-constant exogenous regressors, we choose $Q(x_1, x_2) = \Pi_{l=1}^2(2\pi)^{-1/2}\exp\left(-x_l^2/2\right)$.

As it is difficult to specify the optimal bandwidth sequences $h = (h_1, \ldots, h_{p_c})'$ and $\widetilde{h} = (\widetilde{h}_1, \ldots, \widetilde{h}_{p_c})'$ ($p_c = 1$, or 2 here), we propose to choose them via the following two-step procedure. (1) Choose $h = (h_1, \ldots, h_{p_c})'$ and $\lambda = (\lambda_1, \ldots, \lambda_{p_d})'$ with $h_l = s_{X_{nl}^c}n^{-1/3.5}$ and $\lambda_l = s_{X_{nl}^d}n^{-2/3.5}$ to obtain a preliminary estimate $\overrightarrow{\rho}_n$ of ρ_n^0, where $s_{X_{nl}^c}$ denotes the sample standard deviation of X_{nl}^c for $l = 1, \ldots, p_c$, and $s_{X_{nl}^d}$ is similarly defined. (2) Conduct the least squares cross-validation (LSCV) to choose $(\widetilde{h}, \widetilde{\lambda})$ by regressing

Table 1
Parametric estimates and hypothesis tests for DGPs 1–3.

Estimator	True value	$n = 200$					$n = 800$				
		Mean	Theoret. std dev	Simul. std dev	Sep. test	Joint test	Mean	Theoret. std dev	Simul. std dev	Sep. test	Joint test
DGP 1											
$\widetilde{\rho}_n$	0	−0.034	0.117	0.135	0.067		−0.016	0.058	0.061	0.064	
$\widetilde{\gamma}_n$	0	0.042	0.155	0.147	0.049	0.072	0.019	0.074	0.074	0.054	0.064
$\widehat{\rho}_{KP}$	0	0.008	0.182	0.192	0.068		0.004	0.093	0.096	0.050	
$\widehat{\gamma}_{KP}$	0	−0.013	0.201	0.205	0.057	0.075	0.002	0.104	0.109	0.074	0.048
$\widetilde{\rho}_n$	0.3	0.268	0.112	0.131	0.689		0.284	0.055	0.058	0.996	
$\widetilde{\gamma}_n$	0	0.043	0.150	0.149	0.059	0.955	0.020	0.075	0.075	0.060	1
$\widehat{\rho}_{KP}$	0.3	0.310	0.172	0.182	0.483		0.297	0.088	0.091	0.900	
$\widehat{\gamma}_{KP}$	0	−0.024	0.201	0.204	0.062	0.928	−0.001	0.103	0.109	0.078	1
$\widetilde{\rho}_n$	0	−0.035	0.129	0.149	0.063		−0.012	0.064	0.068	0.054	
$\widetilde{\gamma}_n$	0.3	0.294	0.157	0.141	0.539	0.792	0.300	0.074	0.074	0.982	1
$\widehat{\rho}_{KP}$	0	0.013	0.185	0.196	0.070		0.007	0.095	0.102	0.058	
$\widehat{\gamma}_{KP}$	0.3	0.085	0.202	0.210	0.089	0.178	0.096	0.104	0.111	0.158	0.538
$\widetilde{\rho}_n$	0.3	0.266	0.125	0.145	0.633		0.288	0.061	0.066	0.970	
$\widetilde{\gamma}_n$	0.3	0.296	0.151	0.144	0.546	0.998	0.300	0.075	0.076	0.980	1
$\widehat{\rho}_{KP}$	0.3	0.315	0.176	0.188	0.494		0.308	0.090	0.098	0.864	
$\widehat{\gamma}_{KP}$	0.3	0.073	0.204	0.209	0.080	0.994	0.093	0.105	0.112	0.148	1
DGP 2											
$\widetilde{\rho}_n$	0	−0.007	0.077	0.085	0.074		−0.006	0.039	0.041	0.068	
$\widetilde{\gamma}_n$	0	0.005	0.122	0.125	0.061	0.067	0.003	0.061	0.062	0.054	0.058
$\widehat{\rho}_{KP}$	0	0.005	0.093	0.100	0.079		−0.002	0.048	0.049	0.042	
$\widehat{\gamma}_{KP}$	0	−0.012	0.131	0.134	0.058	0.072	−0.001	0.066	0.069	0.062	0.050
$\widetilde{\rho}_n$	0.3	0.294	0.073	0.081	0.953		0.296	0.037	0.040	1	
$\widetilde{\gamma}_n$	0	0.005	0.122	0.126	0.066	0.990	0.006	0.061	0.067	0.070	1
$\widehat{\rho}_{KP}$	0.3	0.305	0.088	0.094	0.897		0.297	0.046	0.047	0.998	
$\widehat{\gamma}_{KP}$	0	−0.014	0.131	0.134	0.058	0.975	−0.001	0.067	0.066	0.040	1
$\widetilde{\rho}_n$	0	0.008	0.083	0.092	0.073		−0.008	0.043	0.047	0.064	
$\widetilde{\gamma}_n$	0.3	0.279	0.119	0.122	0.624	0.781	0.295	0.060	0.066	0.992	1
$\widehat{\rho}_{KP}$	0	0.007	0.095	0.103	0.081		−0.002	0.050	0.052	0.062	
$\widehat{\gamma}_{KP}$	0.3	0.160	0.129	0.137	0.269	0.421	0.171	0.067	0.069	0.714	0.924
$\widetilde{\rho}_n$	0.3	0.293	0.079	0.088	0.897		0.295	0.040	0.041	1	
$\widetilde{\gamma}_n$	0.3	0.270	0.119	0.123	0.619	1	0.291	0.060	0.060	0.996	1
$\widehat{\rho}_{KP}$	0.3	0.307	0.092	0.100	0.872		0.300	0.048	0.047	1	
$\widehat{\gamma}_{KP}$	0.3	0.156	0.131	0.138	0.253	1	0.167	0.067	0.069	0.690	1
DGP 3											
$\widetilde{\rho}_n$	0	0.005	0.078	0.090	0.071		−0.002	0.036	0.038	0.058	
$\widetilde{\gamma}_n$	0	−0.010	0.122	0.126	0.059	0.066	−0.001	0.060	0.063	0.050	0.060
$\widehat{\rho}_{KP}$	0	0.007	0.090	0.095	0.056		0.001	0.045	0.043	0.044	
$\widehat{\gamma}_{KP}$	0	−0.018	0.130	0.130	0.063	0.053	−0.001	0.065	0.065	0.054	0.036
$\widetilde{\rho}_n$	0.3	0.305	0.075	0.085	0.943		0.298	0.035	0.036	1	
$\widetilde{\gamma}_n$	0	−0.011	0.123	0.126	0.057	0.989	−0.001	0.060	0.063	0.050	1
$\widehat{\rho}_{KP}$	0.3	0.309	0.085	0.090	0.915		0.301	0.042	0.040	1	
$\widehat{\gamma}_{KP}$	0	−0.022	0.129	0.130	0.062	0.982	−0.002	0.065	0.062	0.044	1
$\widetilde{\rho}_n$	0	0.003	0.083	0.099	0.074		−0.000	0.039	0.043	0.058	
$\widetilde{\gamma}_n$	0.3	0.259	0.120	0.127	0.655	0.839	0.287	0.057	0.060	1	1
$\widehat{\rho}_{KP}$	0	0.004	0.091	0.097	0.060		−0.000	0.046	0.045	0.048	
$\widehat{\gamma}_{KP}$	0.3	0.168	0.127	0.130	0.280	0.410	0.174	0.064	0.065	0.780	0.924
$\widetilde{\rho}_n$	0.3	0.306	0.079	0.092	0.907		0.300	0.037	0.040	1	
$\widetilde{\gamma}_n$	0.3	0.249	0.120	0.124	0.630	1	0.286	0.058	0.061	1	1
$\widehat{\rho}_{KP}$	0.3	0.312	0.088	0.096	0.896		0.300	0.044	0.043	1	
$\widehat{\gamma}_{KP}$	0.3	0.148	0.129	0.132	0.228	0.999	0.173	0.064	0.065	0.774	1

$Y_n - \overrightarrow{\rho}_n W_n Y_n$ on X_n. Since undersmoothing is required for h and λ, we set $h = \widetilde{h} n^{-1/3.5+1/(p_c+4)}$ and $\lambda = \widetilde{\lambda} n^{-2/3.5+2/(p_c+4)}$ and obtain an updated estimator $\widetilde{\rho}_n$ of ρ_n^0.

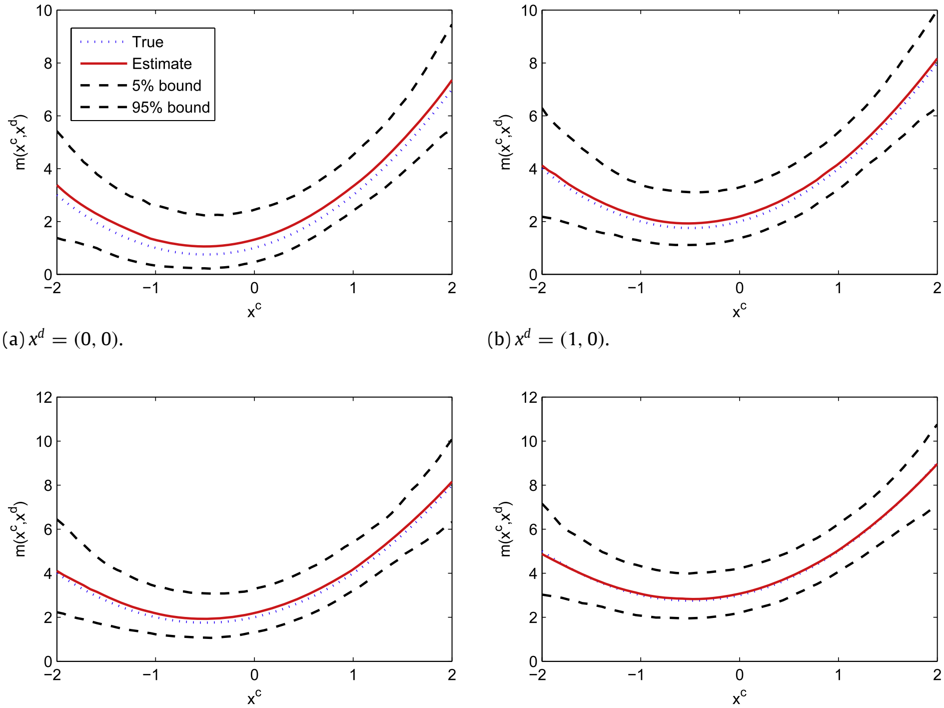
To be concrete, let $\overrightarrow{y}_{n,i}$ denote the ith element of $Y_n - \overrightarrow{\rho}_n W_n Y_n$. Let $\overrightarrow{m}_{-i}\left(x_{n,i}\right)$ be the leave-one-out local linear estimator of $m\left(x_{n,i}\right)$ by leaving the observation $(x_{n,i}, \overrightarrow{y}_{n,i})$ out in the estimation procedure and by using the smoothing parameters (h, λ). We choose $(\widetilde{h}, \widetilde{\lambda})$ to be

$$(\widetilde{h}, \widetilde{\lambda}) = \arg\min_{(h,\lambda)} n^{-1} \sum_{i=1}^{n} \left[\overrightarrow{y}_{n,i} - \overrightarrow{m}_{-i}\left(x_{n,i}\right)\right]^2 w(x_{n,i}^c),$$

where $w(x_{n,i}^c)$ is a nonnegative weight function. In our simulations, we set $w(x_{n,i}^c) = \Pi_{s=1}^{p_c} 1(|x_{n,is}^c - \overline{x}_s^c|) \leq 2 s_{X_{ns}^c}$ with $\overline{x}_s^c$ being the sample mean of X_{ns}^c.

After obtaining $\widetilde{\rho}_n$, we obtain the estimate $\widetilde{m}\left(x_{n,i}\right)$ of $m\left(x_{n,i}\right)$ as in Section 2 and the estimate $\widetilde{\gamma}_n$ of γ_n^0 as in Section 4.

Recently, Kelejian and Prucha (2010) study the generalized moments (GM) estimation of spatial autoregressive models with autoregressive and heteroscedastic errors. Note that their estimation procedure requires a complete specification of the regression model, i.e., the functional form of $m\left(x_{n,i}\right)$ has to be known and is actually linear in their paper. In the case of nonlinearity, it is hard to



(a) $x^d = (0, 0)$.

(b) $x^d = (1, 0)$.

(c) $x^d = (0, 1)$.

(d) $x^d = (1, 1)$.

Fig. 1. Estimation of $m(x^c, x^d)$ in DGP 1 with $(\rho_n^0, \gamma_n^0) = (0.3, 0.3)$.

know the exact form of $m\left(x_{n,i}\right)$. To check the robustness of Kelejian and Prucha's (2010) procedure against nonlinearity, we will also report their GM estimators of (ρ_n^0, γ_n^0) by pretending that $m\left(x_{n,i}\right)$ is linear in $x_{n,i}$ in all DGPs. We denote their estimators of ρ_n^0 and γ_n^0 as $\widehat{\rho}_{KP}$ and $\widehat{\gamma}_{KP}$, respectively.

For each Monte Carlo experiment, we consider samples of size $n = 200$ and 800. The numbers of Monte Carlo replications are 1000 and 500 for the cases $n = 200$ and $n = 800$, respectively.

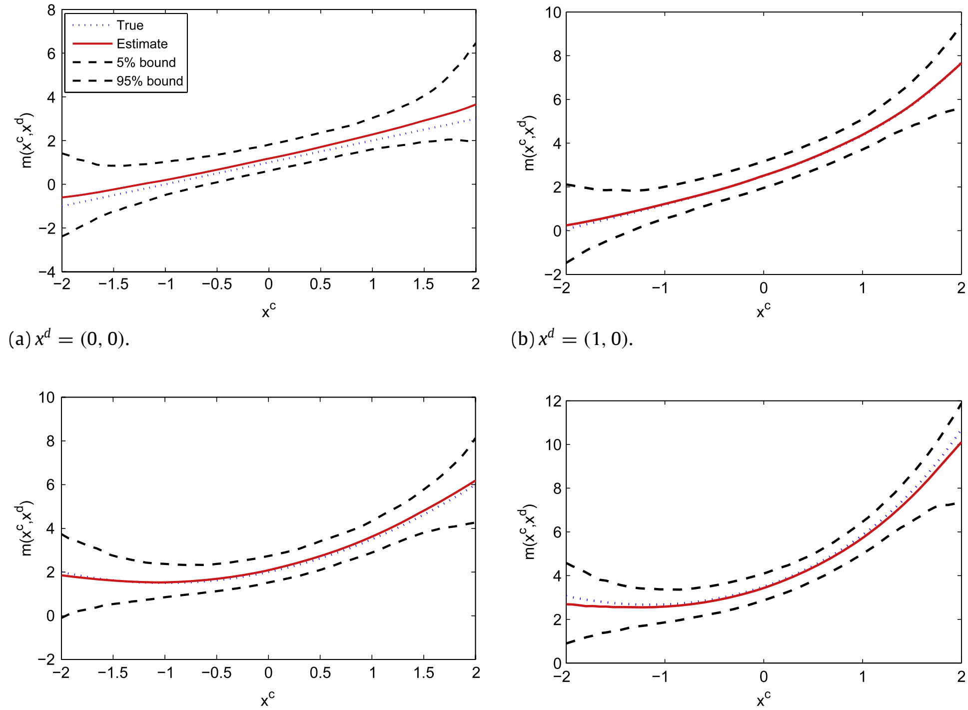
Table 1 reports the empirical mean, theoretical standard deviations (theoret. std dev), simulated standard deviations (simul. std dev), separate null hypothesis test (sep. test), and joint hypothesis test (joint test) results regarding the parametric component in the model. The theoretical standard deviations are calculated from the asymptotic variance-covariance formula and then averaged over the 1000 or 500 replications; the simulated standard deviations are the empirical standard deviations of the corresponding estimators obtained in the replications. The separate tests are conducted to test whether each spatial parameter is significantly different from zero in each replication; the joint tests are to test whether the two spatial parameters are jointly zero in each replication. Both tests are performed at a nominal 5% significance level.

We summarize some important findings from Table 1. First, we see that the estimates $\widetilde{\rho}_n$ are well behaved in all cases and the estimates $\widetilde{\gamma}_n$ are well behaved for all scenarios in DGPs 1 and 2. For DGP 3, we observe non-negligible small sample biases when $n = 200$ but the biases diminish rapidly as the sample size quadruples. Second, despite the use of the linear approximation of $m(\cdot)$, the estimates $\widehat{\rho}_{KP}$ behave quite well in all cases under our investigation. In contrast, the estimates $\widehat{\gamma}_{KP}$ can be seriously biased downwards. For example, when $(\rho_n^0, \gamma_n^0) = (0.3, 0.3)$ in DGP 1, the average values of $\widehat{\gamma}_{KP}$ are 0.073 and 0.093 for $n = 200$ and 800, respectively. Third, the simulated standard deviations are largely consistent with theoretical standard deviations in all cases. As the sample size quadruples, we observe that both sets of standard deviations are roughly halved as predicted by the theory. Fourth, the rejection rates of the separate and joint tests are reasonably close to the nominal level 5% when the corresponding null hypothesis is true. When the null hypotheses are not true, we observe larger values of rejection rates for the tests based upon our semiparametric estimates $(\widetilde{\rho}_n, \widetilde{\gamma}_n)$ than those based upon Kelejian and Prucha's estimates $(\widehat{\rho}_{KP}, \widehat{\gamma}_{KP})$. This is as expected because $\widehat{\gamma}_{KP}$ tends to be biased downwards. For example, consider testing $H_0 : \gamma_n^0 = 0$ in DGP 2 when the true value of (ρ_n^0, γ_n^0) is (0.3, 0.3). The rejection rates for our test are 0.624 and 1 respectively for $n = 200$ and 800, whereas the rejection rates for Kelejian and Prucha's test are 0.269 and 0.924 respectively for $n = 200$ and 800.

Figs. 1 and 2 plot the estimates $\widetilde{m}(x^c, x^d)$ of $m(x^c, x^d)$ in DGPs 1 and 2, respectively, where $n = 200$, $(\rho_n^0, \gamma_n^0) = (0.3, 0.3)$, and x^d may take four different values: $x^d \equiv (x_1^d, x_2^d) = (0, 0)$, (1, 0), (0, 1), and (1, 1). In each sub-graph of Figs. 1 and 2, we plot the true regression curve $m\left(x^c, x^d\right)$, the median estimate $\widetilde{m}\left(x^c, x^d\right)$ in the 1000 replications, the upper and lower 5% quantiles of the estimates $\widetilde{m}\left(x^c, x^d\right)$ in the replications. In each case, we see the median value of the estimates $\widetilde{m}\left(x^c, x^d\right)$ can trace the true regression curve quite well. As expected, we observe that the variations of $\widetilde{m}\left(x^c, x^d\right)$ become larger and larger as x^c moves from the centered value (0 here) to the tailed values (± 2 here).

For DGP 1, $m\left(x^c, x^d\right) = 1 + x_1^d + x_2^d + x^c\left(1 + x^c\right)$, so there is no interaction between x_1^d, x_2^d and x^c and the true curves $m\left(x^c, x^d\right)$ for different values of x^d parallel to each other. In fact we can observe this phenomenon for $\widetilde{m}\left(x^c, x^d\right)$ in Fig. 1. In contrast, for DGP 2, $m(x^c, x^d) = 1 + x_1^d + x_2^d + x^c + 0.5x_1^d \exp(x^c) + 0.5x_2^d (x^c)^2$, so there are interactions between x_1^d, x_2^d, and x^c, and the shapes of $m(x^c, x^d)$ are quite different from each other for different values of x^d, depending on whether $x_1^d = 0$ or 1 as well as $x_2^d = 0$ or 1. For example, if $x^d \equiv (x_1^d, x_2^d) = (0, 0)$, then $m(x^c, x^d)$ is a linear function of x^c; if $x^d \equiv (x_1^d, x_2^d) = (1, 0)$, then $m(x^c, x^d)$ is an exponentially increasing function of x^c. Fig. 2 indicates the estimates $\widetilde{m}(x^c, x^d)$ can capture such features quite well.

In Fig. 3(a) and (b), we plot the true regression curve $m(x^c, x^d) = 1 + x_1^d + x_2^d + x_1^c + x_2^c + x_1^d \cos\left(0.5\pi x_1^c\right) + x_2^d \sin\left(0.5\pi x_2^c\right) + 0.5x_1^c x_2^c$ in



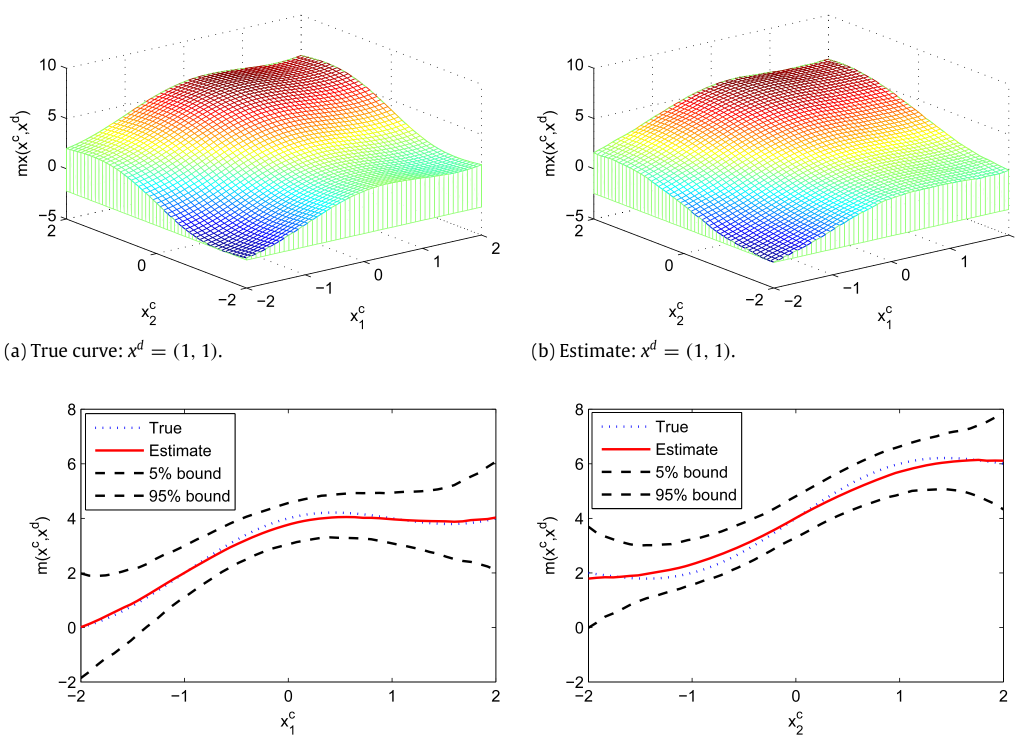
(a) $x^d = (0, 0)$.

(b) $x^d = (1, 0)$.

(c) $x^d = (0, 1)$.

(d) $x^d = (1, 1)$.

Fig. 2. Estimation of $m(x^c, x^d)$ in DGP 2 with $(\rho_n^0, \gamma_n^0) = (0.3, 0.3)$.



(a) True curve: $x^d = (1, 1)$.

(b) Estimate: $x^d = (1, 1)$.

(c) $x_2^c = 0, x^d = (1, 1)$.

(d) $x_1^c = 0, x^d = (1, 1)$.

Fig. 3. Estimation of $m(x^c, x^d)$ in DGP 3 with $(\rho_n^0, \gamma_n^0) = (0.3, 0.3)$. Note: (a) True curve with $x^d = (1, 1)$, (b) Estimated curve with $x^d = (1, 1)$, (c) Projection onto the x_1^c plane, (d) Projection onto the x_2^c plane.

DGP 3 and its estimate $\widetilde{m}(x^c, x^d)$ for the case where $x^d \equiv (x_1^d, x_2^d) = (1, 1)$, $(\rho_n^0, \gamma_n^0) = (0.3, 0.3)$, and $n = 200$. We observe that $\widetilde{m}(x^c, x^d)$ can mimic the shape of $m(x^c, x^d)$ quite well for such small sample size as $n = 200$. Fig. 3(c) plots the median estimate of $m(x^c, x^d)$ versus x_1^c when $(x_2^c, x_1^d, x_2^d) = (0, 1, 1)$, whereas Fig. 3(d) plots the estimate of $m(x^c, x^d)$ versus x_2^c when $(x_1^c, x_1^d, x_2^d) = (0, 1, 1)$. For both subgraphs, we also plot the true regression curve and the upper and lower 5% quantiles of the estimates $\widetilde{m}(x^c, x^d)$ in the replications. We see that in each case, our estimates $\widetilde{m}(x^c, x^d)$ move closely with the true regression curve.

7. Concluding remarks

In this paper we propose semiparametric GMM estimation of SAR models where the error term may exhibit heteroscedasticity or spatial dependence. When the error term follows an SAR process, we also demonstrate that the parameter in the error term can be estimated consistently and one can establish the joint asymptotic distribution for both spatial parameters in the model. Consistent estimation of the asymptotic variance-covariance matrices are also provided. A small set of Monte Carlo simulations are conducted to show the proposed estimators are well behaved in finite samples.

Several extensions are possible. First we conjecture that we can extend our analysis to the case of semiparametric SARAR or SARMA models, where some of the exogenous regressors are parametrically specified in the regression model. Second, after the estimation of the semiparametric SARAR model, we may consider updating the SPGMM estimators proposed in Section 3, say, by considering the Cochrane–Orcutt-type transformed model

$$(I_n - \widetilde{\gamma}_n W_{2n}) Y_n = (I_n - \widetilde{\gamma}_n W_{2n}) \mathbf{m}(X_n) + \rho_n^0 (I_n - \widetilde{\gamma}_n W_{2n}) W_{1n} Y_n + \varepsilon_n.$$

Nevertheless, due to the presence of the nonparametric component $\mathbf{m}$, it is not obvious how one should proceed along this direction (see Xiao et al. (2003) in the time series setup). Third, one can extend the analysis in Section 4.3 by considering the joint GMM estimation of ρ_n^0 and γ_n^0 based on both linear and quadratic moment conditions for $\varepsilon_n = \left(\varepsilon_{n,1}, \ldots, \varepsilon_{n,n}\right)'$ as in Liu et al. (2010). In the special case where the exogenous regressors enter the SARAR model linearly and $\varepsilon_{n,i}$, $i = 1, \ldots, n$, are i.i.d., they consider the optimal GMM estimation of the finite dimensional parameters within the class of linear and quadratic moment conditions. Yet it is not clear whether one can extend their approach to our framework and find the best GMM estimator for the finite dimension parameters ρ_n^0 and γ_n^0 within the class of linear and quadratic moment conditions. We leave these topics for future research.

Appendix

Let C signify a generic constant whose exact value may vary from case to case. Frequently we will use two evident facts (see, e.g., (Kelejian and Prucha, 1999; Lee, 2002)):

Fact 1. If the row and column sums of the $n \times n$ matrices B_{1n} and B_{2n} are uniformly bounded in absolute value, then the row and column sums of $B_{1n}B_{2n}$ are also uniformly bounded in absolute value.

Fact 2. If the row (resp. column) sums of B_{1n} are uniformly bounded in absolute value and B_{2n} is a conformable matrix whose elements are uniformly $O(o_n)$, then so are the elements of $B_{1n}B_{2n}$ (resp. $B_{2n}B_{1n}$).

For example, the row and column sums of $G_{1n} = W_{1n}\left(I_n - \rho_n^0 W_{1n}\right)^{-1}$ are uniformly bounded by Assumption 1 and Fact 1. Noting that the elements $\mathbf{s}_{h\lambda,ij}$ of $\mathbf{S}_{h\lambda}$ are uniformly $O(n^{-1}\Pi_{s=1}^{p_c}\widetilde{h}_s^{-1})$, so are elements of $G_{1n}\mathbf{S}_{h\lambda}$ or $\mathbf{S}_{h\lambda}G_{1n}$ by Assumption 1, Facts 1 and 2.

Appendix A. Some useful results

Here we provide a theorem and a lemma that are used in the proof of the main theorems in the text.

We first consider the linear quadratic forms

$$\mathbf{Q}_{sn} = \varepsilon_n' \mathbf{A}_{sn} \varepsilon_n + \mathbf{a}_{sn}' \varepsilon_n, \quad s = 1, \ldots, r$$

where $\varepsilon_n = (\varepsilon_{n,1}, \ldots, \varepsilon_{n,n})'$ is as defined in Assumption 2, $\mathbf{A}_{sn} = \left(\mathbf{a}_{sn,ij}\right)_{i,j=1,\ldots,n}$ is an $n \times n$ nonstochastic matrix, and $\mathbf{a}_{sn} = \left(\mathbf{a}_{sn,1}, \ldots, \mathbf{a}_{sn,n}\right)'$ is an $n \times 1$ nonstochastic real vector. Let $\mathbf{Q}_n = [\mathbf{Q}_{1n}, \ldots, \mathbf{Q}_{rn}]'$.

Let $\mu_{\mathbf{Q}_{sn}} = E\mathbf{Q}_{sn}$ and $\sigma_{\mathbf{Q}_{stn}} = \operatorname{cov}(\mathbf{Q}_{sn}, \mathbf{Q}_{tn})$ for $s, t = 1, \ldots, r$. Kelejian and Prucha (2010, Lemma A.1) show

$$\mu_{\mathbf{Q}_{sn}} = \sum_{i=1}^{n} \mathbf{a}_{sn,ii}\sigma_{n,i}^2, \tag{A.1}$$

$$\begin{aligned}\sigma_{\mathbf{Q}_{stn}} = {} & 2\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_{sn,ij}\mathbf{a}_{tn,ij}\sigma_{n,i}^2\sigma_{n,j}^2 + \sum_{i=1}^{n} \mathbf{a}_{sn,i}\mathbf{a}_{tn,i}\sigma_{n,i}^2 \\ & + \sum_{i=1}^{n} \mathbf{a}_{sn,ii}\mathbf{a}_{tn,ii}\left[\mu_{n,i}^{(4)} - 3\sigma_{n,i}^4\right] \\ & + \sum_{i=1}^{n} \left(\mathbf{a}_{sn,i}\mathbf{a}_{tn,ii} + \mathbf{a}_{tn,i}\mathbf{a}_{sn,ii}\right)\mu_{n,i}^{(3)},\end{aligned} \tag{A.2}$$

where $\mu_{n,i}^{(k)} = E[\varepsilon_{n,i}^k]$ for $k = 3, 4$. Clearly, if $\mathbf{a}_{sn,ii} = 0$ for all $s = 1, \ldots, r$ and $i = 1, \ldots, n$, $\mu_{\mathbf{Q}_{sn}} = 0$ and the last two terms in the expression of $\sigma_{\mathbf{Q}_{stn}}$ drop out. Let $\boldsymbol{\mu}_{\mathbf{Q}_n} = [\mu_{\mathbf{Q}_{1n}}, \ldots, \mu_{\mathbf{Q}_{rn}}]'$, $\mathbf{\Sigma}_{\mathbf{Q}_n} = \left(\sigma_{\mathbf{Q}_{stn}}\right)_{s,t=1,\ldots,r}$, and $\mathbf{\Sigma}_{\mathbf{Q}_n} = \mathbf{\Sigma}_{\mathbf{Q}_n}^{1/2}(\mathbf{\Sigma}_{\mathbf{Q}_n}^{1/2})'$. The following theorem is proved in Kelejian and Prucha (2010).

Theorem A.1 (*A CLT for Linear Quadratic Forms*). *Suppose that for $s = 1, 2, \ldots, r$, $\mathbf{A}_{sn}$ is symmetric and the row (column) sums of $\mathbf{A}_{sn}$ are uniformly bounded. Suppose that $\sup_n n^{-1}\sum_{i=1}^n \left|\mathbf{a}_{sn,i}\right|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$. Suppose that $n^{-1}\lambda_{\min}\left(\mathbf{\Sigma}_{\mathbf{Q}_n}\right) \geq c$ for some $c > 0$. Then*

$$\mathbf{\Sigma}_{\mathbf{Q}_n}^{-1/2}\left(\mathbf{Q}_n - \boldsymbol{\mu}_{\mathbf{Q}_n}\right) \xrightarrow{d} N\left(0, I_r\right),$$

where elements of $\boldsymbol{\mu}_{\mathbf{Q}_n}$ and $\mathbf{\Sigma}_{\mathbf{Q}_n}$ are given by (A.1) and (A.2), respectively.

Recall that $\mathbf{S}_{h\lambda} = \left(\mathbf{s}_{h\lambda}\left(x_{n,1}\right), \ldots, \mathbf{s}_{h\lambda}\left(x_{n,n}\right)\right)'$. Denote the (i,j)th element of $\mathbf{S}_{h\lambda}$ as $\mathbf{s}_{h\lambda,ij}$. To study the properties of $\mathbf{s}_{h\lambda,ij}$, we need to distinguish whether $x_{n,j}^c \equiv (x_{n,j1}^c, \ldots, x_{n,jp_c}^c)'$ is a boundary point in the compact support $\mathcal{X}^c$ of $f(x^c, x^d)$. Without loss of generality, we assume $\mathcal{X}^c = \Pi_{s=1}^{p_c} S_s$, where $S_s \equiv [\underline{x}_s, \overline{x}_s]$. A point $x_{n,j}^c$ is said to be a boundary point in $\mathcal{X}^c$ if there exists $s \in \{1, 2, \ldots, p_c\}$ such that $x_{n,js}^c = \underline{x}_s + b_s h_s$ or $x_{n,js}^c = \overline{x}_s - c_s h_s$ for some finite positive numbers b_s and c_s. Otherwise, we say that $x_{n,j}^c$ is not a boundary point. In the following, when $x_{n,j}^c$ is a boundary point, we assume that it is a pure "lower" boundary point such that we can write $x_{n,j}^c = \underline{x} + b \odot h_s$, where $\underline{x} = (\underline{x}_1, \ldots \underline{x}_{p_c})'$ and $b = (b_1, \ldots, b_{p_c})'$. Other cases of boundary points can be analogously analyzed.

Define

$$\mathbf{A}_b(x) = \begin{pmatrix} \kappa_{b,0}\varphi(x) & \varphi(x)\kappa_{b,1}' \\ \varphi(x) \otimes \kappa_{b,1} & \varphi(x) \otimes \kappa_{b,2} \end{pmatrix}$$

where $\kappa_{b,0} = \int_{-b}^{\infty} \Pi_{s=1}^{p_c} q(u_s)\, du$, $\kappa_{b,1} = \int_{-b}^{\infty} u\Pi_{s=1}^{p_c} q(u_s)\, du$, and $\kappa_{b,2} = \int_{-b}^{\infty} uu'\Pi_{s=1}^{p_c} q(u_s)\, du$. Note that when $b = \infty$, $\Pi_b = \Pi_\infty = \Pi$, where Π is defined in Section 3.3.

Lemma A.2. (a) $\sum_{j=1}^n \mathbf{s}_{h\lambda,ij} = 1$ *and* $\sum_{j=1}^n \mathbf{s}_{h\lambda,ij}\left(x_{n,j}^c - x_{n,i}^c\right) = 0_{p_c \times 1}$ *for each i;*

(b) $\mathbf{A}_{n,h\lambda}\left(x_{n,j}\right) = f\left(x_{n,j}\right)\overline{\mathbf{A}}\left(x_{n,j}\right) + o(1)$ *for each j;*

(c) *the row and column sums of $\mathbf{S}_{h\lambda} = (\mathbf{s}_{h\lambda,ij})$ are uniformly bounded in absolute value for sufficiently large n,*

where $\overline{\mathbf{A}} = \mathbf{A}$ if $x_{n,j}$ is not a boundary point, and $\overline{\mathbf{A}} = \mathbf{A}_b$ if $x_{n,j}^c = \underline{x} + b \odot h$. For other cases of boundary points, $\overline{\mathbf{A}}$ can be similarly defined.

The proof of the above lemma and other lemmata in this paper can be found at http://www.mysmu.edu/faculty/ljsu/Publications/Spatial_NPGMM_supplement.pdf.

Appendix B. Proof of results in Section 3

Proof of Theorem 3.1. Noting that $Y_n = \rho_n^0 W_{1n} Y_n + \mathbf{m}(X_n) + U_n$, we have

$$\sqrt{n}\left(\widetilde{\rho}_n - \rho_n^0\right) = \frac{(n^{-1}\widetilde{\widetilde{Y}}_n' Z_n)\Omega_n[n^{-1/2}Z_n'(I_n - \mathbf{S}_{h\lambda})U_n]}{(n^{-1}\widetilde{\widetilde{Y}}_n' Z_n)\Omega_n(n^{-1}Z_n'\widetilde{\widetilde{Y}}_n)} + \frac{(n^{-1}\widetilde{\widetilde{Y}}_n' Z_n)\Omega_n[n^{-1/2}Z_n'(I_n - \mathbf{S}_{h\lambda})\mathbf{m}(X_n)]}{(n^{-1}\widetilde{\widetilde{Y}}_n' Z_n)\Omega_n(n^{-1}Z_n'\widetilde{\widetilde{Y}}_n)}.$$

Noting that $n^{-1}Z_n'\widetilde{\widetilde{Y}}_n = B + o_p(1)$ and $\Omega_n = \Omega + o_p(1)$ by Assumption 5, it suffices to prove the theorem by showing that

$$T_{n1} \equiv n^{-1/2}Z_n'(I_n - \mathbf{S}_{h\lambda})\mathbf{m}(X_n) = o(1) \tag{B.1}$$

and

$$T_{n2} \equiv n^{-1/2}Z_n'(I_n - \mathbf{S}_{h\lambda})U_n \xrightarrow{d} N(0, \Theta), \tag{B.2}$$

because then $\sqrt{n}\left(\widetilde{\rho}_n - \rho_n^0\right) = \frac{B'\Omega[n^{-1/2}Z_n'(I_n - \mathbf{S}_{h\lambda})U_n]}{B'\Omega B} + o_p(1) \xrightarrow{d} N\left(0, \left(B'\Omega B\right)^{-2} B'\Omega\Theta\Omega B\right)$.

We first show (B.1). Recall $\dot{m}(x) = \partial m(x)/\partial x^c$. Let $\ddot{m}(x) = \partial^2 m(x)/\partial x^c \partial x^{c\prime}$. Then $m(x_{n,i}) - m(x_{n,j}) = \dot{m}(x_{n,j})'\left((x_{n,i}^c - x_{n,j}^c)\right) + \frac{1}{2}\left(x_{n,i}^c - x_{n,j}^c\right)'\ddot{m}(x_{n,j})\left(x_{n,i}^c - x_{n,j}^c\right) + o(\|h\|^2)$ if $\left\|x_{n,i}^c - x_{n,j}^c\right\| \le C\|h\|$ and $x_{n,i}^d = x_{n,j}^d$. By Lemma A.2(a),

$$\begin{aligned} T_{n1} &= n^{-1/2}\sum_{i=1}^n z_{n,i}\left[m(x_{n,i}) - \sum_{j=1}^n \mathbf{s}_{h\lambda,ij} m(x_{n,j})\right] \\ &= n^{-1/2}\sum_{i=1}^n z_{n,i}\sum_{j=1}^n \mathbf{s}_{h\lambda,ij}\left[m(x_{n,i}) - m(x_{n,j})\right] \\ &= T_{n11} + T_{n12} + o\left(n^{1/2}\|h\|^2\right), \end{aligned}$$

where

$$\begin{aligned} T_{n11} &= \frac{n^{-1/2}}{2}\sum_{i=1}^n z_{n,i}\sum_{j=1}^n \mathbf{s}_{h\lambda,ij}\frac{1}{2}\left(x_{n,i}^c - x_{n,j}^c\right)'\ddot{m}(x_{n,j}) \\ &\quad \times \left(x_{n,i}^c - x_{n,j}^c\right)\mathbf{1}\left(x_{n,i}^d = x_{n,j}^d\right) \quad \text{and} \\ T_{n12} &= n^{-1/2}\sum_{i=1}^n z_{n,i}\sum_{j=1}^n \mathbf{s}_{h\lambda,ij} \\ &\quad \times \left[m(x_{n,i}) - m(x_{n,j})\right]\mathbf{1}\left(x_{n,i}^d \ne x_{n,j}^d\right). \end{aligned}$$

As in the proof of Lemma A.2(c), let $e_1'\left(\overline{\mathbf{A}}(x)'\overline{\mathbf{A}}(x)\right)^{-1}\overline{\mathbf{A}}(x)' = (a_1(x)', a_2(x)')$. By Assumptions 3, 4 and Lemma A.2(b),

$$\begin{aligned} T_{n11} &= \frac{n^{-1/2}}{2}\sum_{i=1}^n z_{n,i} f^{-1}(x_{n,i})\, n^{-1} \\ &\quad \times \sum_{j=1}^n \left[a_1(x_{n,i})' z_{n,j}^{(1)} + a_2(x_{n,i})'\left(z_{n,j}^{(1)} \otimes ((x_{n,j}^c - x_{n,i}^c)/h)\right)\right] \\ &\quad \times K_{h\lambda,ij}\left(x_{n,i}^c - x_{n,j}^c\right)'\ddot{m}(x_{n,j})\left(x_{n,i}^c - x_{n,j}^c\right) \\ &\quad \times \mathbf{1}\left(x_{n,i}^d = x_{n,j}^d\right)\{1 + o(1)\} \\ &= \frac{n^{-1/2}}{2}\sum_{i=1}^n z_{n,i} f^{-1}(x_{n,i}) \\ &\quad \times \int \left[a_1(x_{n,i})'\varphi_n(x_{n,i}) + a_1(x_{n,i})'\left(\varphi_n(x_{n,i}) \otimes u\right)\right] \\ &\quad \times \prod_{t=1}^{p_c} q(u_t)\,(h \odot u)'\,\ddot{m}\left(x_{n,i}^c + h \odot u, x_i^d\right)(h \odot u) \\ &\quad \times f\left(x_{n,i}^c + h \odot u, x_{n,i}^d\right) du\,\{1 + o(1)\} \\ &= O\left(n^{1/2}\|h\|^2\right) = o(1). \end{aligned}$$

Similarly, one can show that $T_{n12} = O\left(n^{1/2}\|\lambda\|\right) = o(1)$. Hence $T_{n1} = o(1)$.

We now show (B.2). Let c be an arbitrary $q \times 1$ vector with $\|c\| = 1$. By the Cramér–Wold device, it suffices to show that $c'T_{n2} \xrightarrow{d} N\left(0, c'\Theta c\right)$. Clearly, $E\left[c'T_{n2}\right] = 0$. Let $s_n^2 \equiv E[c'T_{n2}]^2$ and $\widetilde{T}_{n2} = c'T_{n2}/s_n$. Then by construction, $E(\widetilde{T}_{n2}) = 0$ and $E(\widetilde{T}_{n2})^2 = 1$. Write

$$\begin{aligned} \widetilde{T}_{n2} &= n^{-1/2}\sum_{l=1}^n\sum_{j=1}^n\left[c'z_{n,j} - \sum_{i=1}^n c'z_{n,i}\mathbf{s}_{h\lambda,ij}\right]a_{n,jl}\varepsilon_{n,l}/s_n \\ &= \sum_{l=1}^n \widetilde{\varepsilon}_{n,l}, \end{aligned}$$

where $\widetilde{\varepsilon}_{n,l} \equiv \varepsilon_{n,l} n^{-1/2}\sum_{j=1}^n\left[c'z_{n,j} - \sum_{i=1}^n c'z_{n,i}\mathbf{s}_{h\lambda,ij}\right]a_{n,jl}/s_n$. By the triangle inequality, Assumption 5, and Lemma A.2(c), we have that for sufficiently large n,

$$\begin{aligned} \left|\sum_{i=1}^n c'z_{n,i}\mathbf{s}_{h\lambda,ij}\right| &\le \sum_{i=1}^n \left|c'z_{n,i}\mathbf{s}_{h\lambda,ij}\right| \\ &\le c_z\sum_{i=1}^n\left|\mathbf{s}_{h\lambda,ij}\right| = c_z C, \end{aligned} \tag{B.3}$$

where $c_z = \sup_{1\le i\le n, n\ge 1}\left\|z_{n,i}\right\|$. That is, for each $j = 1, \ldots, n$, $\left|\sum_{i=1}^n c'z_{n,i}\mathbf{s}_{h\lambda,ij}\right|$ is bounded by a constant for sufficiently large n. Hence by Assumption 3 and the C_r inequality (e.g., Pagan and Ullah (1999, p. 350)), for some small $\delta > 0$

$$\begin{aligned} \sum_{l=1}^n E\left|\widetilde{\varepsilon}_{n,l}\right|^{2+\delta} &= \frac{n^{-(2+\delta)/2}}{s_n^{2+\delta}} \\ &\quad \times \sum_{l=1}^n\left|\sum_{j=1}^n\left[c'z_{n,j} - \sum_{i=1}^n c'z_{n,i}\mathbf{s}_{h\lambda,ij}\right]a_{n,jl}\right|^{2+\delta} E\left|\varepsilon_{n,l}\right|^{2+\delta} \\ &\le \frac{2^{2+\delta}n^{-(2+\delta)/2}}{s_n^{2+\delta}}\sum_{l=1}^n\left|\sum_{j=1}^n c'z_{n,j}a_{n,jl}\right|^{2+\delta} \\ &\quad + \frac{2^{2+\delta}n^{-(2+\delta)/2}}{s_n^{2+\delta}}\sum_{l=1}^n\left|\sum_{j=1}^n\sum_{i=1}^n c'z_{n,i}\mathbf{s}_{h\lambda,ij}a_{n,jl}\right|^{2+\delta} \\ &\equiv S_{n1} + S_{n2}, \quad \text{say.} \end{aligned}$$

By Assumptions 2 and 5, $S_{n1} \le \frac{Cn^{-(2+\delta)/2}}{s_n^{2+\delta}}\sum_{l=1}^n\left|\sum_{j=1}^n\left|a_{n,jl}\right|\right|^{2+\delta} = O(n^{-\delta/2}) = o(1)$. Similarly, by (B.3) and Assumptions 2 and 5, for sufficiently large n, $S_{n2} \le \frac{Cn^{-(2+\delta)/2}}{s_n^{2+\delta}}\sum_{l=1}^n\left|\sum_{j=1}^n\left|a_{n,jl}\right|\right|^{2+\delta} = O(n^{-\delta/2}) = o(1)$. Hence $\sum_{l=1}^n E\left|\widetilde{\varepsilon}_{n,l}\right|^{2+\delta} = o(1)$. It follows from Theorems 23.6 and 23.11 of Davidson (1994) that $\widetilde{T}_{n2} \xrightarrow{d} N(0, 1)$. The result then follows from the fact that $s_n^2 = n^{-1}c'Z_n'(I_n - \mathbf{S}_{h\lambda})A_n\Sigma_n A_n'(I_n - \mathbf{S}_{h\lambda})'Z_n c \to c'\Theta c$ by Assumption 5(iii). □

Proof of Theorem 3.2. By definition, $\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x) = \widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}(x)'\left(Y_n - \widetilde{\rho}_n W_{1n} Y_n\right) = \widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}(x)'\left(\mathbf{m}(X_n) + U_n + \left(\rho_n^0 - \widetilde{\rho}_n\right) W_{1n} Y_n\right)$, where $\widetilde{\mathbf{s}}_{h\lambda}(x)' \equiv n^{-1}\left(\mathbf{A}_{n,h\lambda}(x)'\mathbf{A}_{n,h\lambda}(x)\right)^{-1}\mathbf{A}_{n,h\lambda}(x)'\mathbf{Z}_{n,h}(x)'\operatorname{diag}\left(\mathbf{k}_{h\lambda}(x)\right)$. It follows that $\sqrt{n\Pi_{s=1}^{p_c}\widetilde{h}_s}\left(\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x) - \mathbf{M}_{\widetilde{h}}(x)\right) = N_{n1} + N_{n2} + N_{n3}$, where

$$N_{n1} = \sqrt{n\prod_{s=1}^{p_c}\widetilde{h}_s}\left(\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}(x)'\,\mathbf{m}(X_n) - \mathbf{M}_{\widetilde{h}}(x)\right),$$

$$N_{n2} = \sqrt{n\prod_{s=1}^{p_c}\widetilde{h}_s}\,\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}(x)'\,U_n, \quad \text{and}$$

$$N_{n3} = \left(\rho_n^0 - \widetilde{\rho}_n\right)\sqrt{n\prod_{s=1}^{p_c}\widetilde{h}_s}\,\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}(x)'\,W_{1n}Y_n.$$

By Theorem 3.1, $\widetilde{\rho}_n - \rho_n^0 = O_p(n^{-1/2})$. With this, it is easy to show $N_{n3} = O_p\left(\sqrt{\Pi_{s=1}^{p_c}\widetilde{h}_s}\right) = o_p(1)$. We now show that N_{n1} and N_{n2} contribute to the asymptotic bias and variance of $\widetilde{\mathbf{M}}_{\widetilde{h}\widetilde{\lambda}}(x)$, respectively.

By the second order Taylor expression,

$$m\left(x_{n,i}\right) = (1, ((x_{n,i}^c - x^c)/\widetilde{h})')\mathbf{M}_{\widetilde{h}}(x) + \frac{1}{2}\left(x_{n,i}^c - x^c\right)'\ddot{m}(x)\left(x_{n,i}^c - x^c\right) + o(\|\widetilde{h}\|^2) \tag{B.4}$$

for $\left\|x_{n,i}^c - x^c\right\| \le C\|\widetilde{h}\|$ and $x_{n,i}^d = x^d$. Denote $\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}\left(x_{n,i}, x\right)$ as a typical column of $\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}(x)'$, i.e., $\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}(x)' = (\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}\left(x_{n,1}, x\right), \ldots, \widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}\left(x_{n,n}, x\right))$. Noting that

$$\begin{aligned} I_{p_c+1} &= \left(\mathbf{A}_{n,\widetilde{h}\widetilde{\lambda}}(x)'\mathbf{A}_{n,\widetilde{h}\widetilde{\lambda}}(x)\right)^{-1}\mathbf{A}_{n,\widetilde{h}\widetilde{\lambda}}(x)'\mathbf{A}_{n,\widetilde{h}\widetilde{\lambda}}(x) \\ &= \sum_{j=1}^{n}\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}\left(x_{n,j}, x\right)\left(1, ((x_{n,j}^c - x)/\widetilde{h})'\right), \end{aligned}$$

we have $N_{n1} = N_{n11} + N_{n12} + o_p(1)$, where

$$\begin{aligned} N_{n11} &= \frac{1}{2}\sqrt{n\prod_{s=1}^{p_c}\widetilde{h}_s}\sum_{i=1}^{n}\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}\left(x_{n,i}, x\right) \\ &\quad \times \left(x_{n,i}^c - x^c\right)'\ddot{m}(x)\left(x_{n,i}^c - x^c\right)\mathbf{1}\left(x_{n,i}^d = x^d\right), \quad \text{and} \\ N_{n12} &= \sqrt{n\prod_{s=1}^{p_c}\widetilde{h}_s}\sum_{i=1}^{n}\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}\left(x_{n,i}, x\right) \\ &\quad \times \left[m\left(x_{n,i}\right) - m(x)\right]\mathbf{1}\left(x_{n,i}^d \neq x^d\right). \end{aligned}$$

Recall $\mathbf{A}^*(x) = \left(\mathbf{A}(x)'\mathbf{A}(x)\right)^{-1}\mathbf{A}(x)'$. Let $\nu_n = \sqrt{n\Pi_{s=1}^{p_c}\widetilde{h}_s}$ and $q_{\widetilde{h}i,x} = \Pi_{t=1}^{p_c}q_{\widetilde{h}it,x}$, where $q_{\widetilde{h}it,x} \equiv \widetilde{h}_t^{-1}q((x_{n,it}^c - x_t^c)/\widetilde{h}_t)$ for $t = 1, 2, \ldots, p_c$, $x_{n,it}^c$ is the tth element of $x_{n,i}^c$, and x_t^c is similarly defined. Then by Lemma A.2(c),

$$\begin{aligned} N_{n11} &= \frac{\{1 + o(1)\}\nu_n}{2}f^{-1}(x)\mathbf{A}^*(x)n^{-1} \\ &\quad \times \sum_{i=1}^{n}\begin{pmatrix} z_{n,i}^{(1)} \\ z_{n,i}^{(1)} \otimes \left(\left(x_{n,i}^c - x^c\right)/\widetilde{h}\right) \end{pmatrix} \\ &\quad \times q_{\widetilde{h}i,x}\left(x_{n,i}^c - x^c\right)'\ddot{m}(x) \\ &\quad \times \left(x_{n,i}^c - x^c\right)\mathbf{1}\left(x_{n,i}^d = x^d\right) \\ &= \frac{\{1 + o(1)\}\nu_n}{2}f^{-1}(x)\mathbf{A}^*(x) \\ &\quad \times \int\prod_{t=1}^{p_c}\begin{pmatrix} \varphi\left(x^c + \widetilde{h}\odot u, x^d\right) \\ \varphi\left(x^c + \widetilde{h}\odot u, x^d\right)\otimes u \end{pmatrix} \\ &\quad \times q(u_t)f(x^c + \widetilde{h}\odot u, x^d) \\ &\quad \times (\widetilde{h}\odot u)\ddot{m}(x)(\widetilde{h}\odot u)du \\ &= \frac{\nu_n}{2}\mathbf{A}^*(x)\begin{pmatrix} \kappa_{21}\varphi(x)\sum_{s=1}^{p_c}\widetilde{h}_s^2 m_{ss}(x) \\ \mathbf{0}_{q_1p_c\times 1} \end{pmatrix} + o(\nu_n\|\widetilde{h}\|^2). \end{aligned}$$

Similarly, N_{n12} is given in Box III, where $\mathbf{1}_s(x_{n,i}^d, x^d)$ is defined in (3.3). So by Assumption 4,

$$N_{n12} = \begin{pmatrix} \sqrt{n\prod_{s=1}^{p_c}\widetilde{h}_s}\,b_{\widetilde{\lambda}}(x) \\ \mathbf{0}_{q_1p_c\times 1} \end{pmatrix} + o(1),$$

where $b_{\widetilde{\lambda}}(x)$ is defined in (3.4). Hence

$$\begin{aligned} N_{n1} &= \sqrt{n\prod_{s=1}^{p_c}\widetilde{h}_s}\,\mathbf{A}^*(x) \\ &\quad \times \begin{pmatrix} \frac{1}{2}\kappa_{21}\varphi(x)\sum_{s=1}^{p_c}\widetilde{h}_s^2 m_{ss}(x) + b_{\widetilde{\lambda}}(x) \\ \mathbf{0}_{q_1p_c\times 1} \end{pmatrix} + o(1). \end{aligned}$$

Note that $N_{n2} = (f^{-1}(x)\mathbf{A}^*(x) + o(1))n^{-1/2}\sqrt{\Pi_{s=1}^{p_c}\widetilde{h}_s}\mathbf{Z}_{n,\widetilde{h}}(x)'\operatorname{diag}\left(\mathbf{k}_{\widetilde{h}\widetilde{\lambda}}(x)\right)U_n + o_p(1)$. Let $c \equiv (c_1', c_2')'$ with $\|c\| = 1$, where c_1 and c_2 are $q_1 \times 1$ and $q_1p_c \times 1$ vectors, respectively. Let

$$\begin{aligned} T_n &\equiv n^{-1/2}\sqrt{\prod_{s=1}^{p_c}\widetilde{h}_s}\,c'\mathbf{Z}_{n,\widetilde{h}}'(x)\operatorname{diag}\left(\mathbf{k}_{\widetilde{h}\widetilde{\lambda}}(x)\right)U_n \\ &= n^{-1/2}\sqrt{\prod_{s=1}^{p_c}\widetilde{h}_s}\sum_{j=1}^{n}\sum_{i=1}^{n}\zeta_{ni}K_{\widetilde{h}\widetilde{\lambda},i}(x)\,a_{n,ij}\varepsilon_{n,j}, \end{aligned}$$

where $\zeta_{ni} \equiv c_1'z_{n,i}^{(1)} + c_2'(z_{n,i}^{(1)} \otimes ((x_{n,i}^c - x^c)/\widetilde{h}))$. Then $E[T_n] = 0$, and

$$\begin{aligned} S_n^2 &\equiv E[T_n]^2 \\ &= n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{l=1}^{n}\zeta_{ni}\zeta_{nl}K_{\widetilde{h}\widetilde{\lambda},i}(x)K_{\widetilde{h}\widetilde{\lambda},l}(x)\,a_{n,ij}a_{n,lj}\sigma_{n,j}^2 \\ &\to c'\Gamma c. \end{aligned}$$

Let $\widetilde{T}_n = T_n/S_n$. We can write $\widetilde{T}_n = \sum_{j=1}^{n}\overline{\varepsilon}_{n,j}$, where

$$\overline{\varepsilon}_{n,j} = n^{-1/2}\sqrt{\prod_{s=1}^{p_c}\widetilde{h}_s}\sum_{i=1}^{n}\zeta_{ni}K_{\widetilde{h}\widetilde{\lambda},i}(x)\,a_{n,ij}\varepsilon_{n,j}/S_n.$$

Let $\widetilde{a}_{n,ij} = \left|a_{n,ij}\right| / \sum_{l=1}^{n}\left|a_{n,lj}\right|$. Then $\widetilde{a}_{n,ij} \ge 0$ for each i and $\sum_{i=1}^{n}\widetilde{a}_{n,ij} = 1$. By Assumptions 2 and 4 and Jensen's inequality, we have that for some small $\delta > 0$,

$$\begin{aligned} \sum_{j=1}^{n}E\left|\overline{\varepsilon}_{n,j}\right|^{2+\delta} &= \frac{n^{-(2+\delta)/2}\left(\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{(2+\delta)/2}}{S_n^{2+\delta}} \\ &\quad \times \sum_{j=1}^{n}E\left|\varepsilon_{n,j}\sum_{i=1}^{n}\zeta_{ni}K_{\widetilde{h}\widetilde{\lambda},i}(x)\,a_{n,ij}\right|^{2+\delta} \end{aligned}$$

$$
\begin{aligned}
N_{n12} &= \{1+o(1)\}\,\nu_n f^{-1}(x)\,\mathbf{A}^*(x)\,n^{-1}\sum_{i=1}^{n}\begin{pmatrix} z_{n,i}^{(1)} \\ z_{n,i}^{(1)}\otimes\left(\left(x_{n,i}^c-x^c\right)/\widetilde{h}\right)\end{pmatrix} K_{\widetilde{h}\widetilde{\lambda},i}(x)\left[m\left(x_{n,i}\right)-m(x)\right]\mathbf{1}\left(x_{n,i}^d\neq x^d\right) \\
&= \{1+o(1)\}\,\nu_n f^{-1}(x)\,\mathbf{A}^*(x)\,n^{-1}\sum_{i=1}^{n}\begin{pmatrix} z_{n,i}^{(1)} \\ z_{n,i}^{(1)}\otimes\left(\left(x_{n,i}^c-x^c\right)/\widetilde{h}\right)\end{pmatrix} q_{\widetilde{h}i,x}\left[m\left(x_{n,i}\right)-m(x)\right]\left\{\sum_{s=1}^{p_d}\widetilde{\lambda}_s\mathbf{1}_s(x_{n,i}^d,x^d)\right\}+O(\|\widetilde{\lambda}\|^2) \\
&= \nu_n f^{-1}(x)\,\mathbf{A}^*(x)\begin{pmatrix}\sum_{v^d\in\mathcal{X}^d}\left[m\left(x^c,v^d\right)-m\left(x^c,x^d\right)\right]\sum_{s=1}^{p_d}\widetilde{\lambda}_s\mathbf{1}_s(v^d,x^d)f\left(x^c,v^d\right)\varphi\left(x^c,v^d\right) \\ \mathbf{0}_{q_1p_c\times 1}\end{pmatrix}+o(\nu_n\|\widetilde{\lambda}\|)
\end{aligned}
$$

Box III.

$$
\begin{aligned}
&\leq \frac{n^{-(2+\delta)/2}\left(\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{(2+\delta)/2}}{S_n^{2+\delta}}\sum_{j=1}^{n}E\left|\varepsilon_{n,j}\right|^{2+\delta} \\
&\quad\times\left|\sum_{l=1}^{n}\left|a_{n,lj}\right|\sum_{i=1}^{n}\left|\zeta_{ni}\right|K_{\widetilde{h}\widetilde{\lambda},i}(x)\,\widetilde{a}_{n,ij}\right|^{2+\delta} \\
&\leq \frac{\overline{\mu}_{2+\delta}n^{-(2+\delta)/2}\left(\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{(2+\delta)/2}}{S_n^{2+\delta}} \\
&\quad\times\sum_{j=1}^{n}\left|\sum_{l=1}^{n}\left|a_{n,lj}\right|\right|^{2+\delta}\sum_{i=1}^{n}\widetilde{a}_{n,ij}\left|\zeta_{ni}\right|^{2+\delta}K_{\widetilde{h}\widetilde{\lambda},i}^{2+\delta}(x) \\
&= \frac{\overline{\mu}_{2+\delta}n^{-(2+\delta)/2}\left(\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{(2+\delta)/2}}{S_n^{2+\delta}} \\
&\quad\times\sum_{j=1}^{n}\left|\sum_{l=1}^{n}\left|a_{n,lj}\right|\right|^{1+\delta}\sum_{i=1}^{n}\left|a_{n,ij}\right|\left|\zeta_{ni}\right|^{2+\delta}K_{\widetilde{h}\widetilde{\lambda},i}^{2+\delta}(x) \\
&\leq \frac{c_a^{1+\delta}\overline{\mu}_{2+\delta}n^{-(2+\delta)/2}\left(\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{(2+\delta)/2}}{S_n^{2+\delta}} \\
&\quad\times\sum_{i=1}^{n}\left|\zeta_{ni}\right|^{2+\delta}K_{\widetilde{h}\widetilde{\lambda},i}^{2+\delta}(x)\sum_{j=1}^{n}\left|a_{n,ij}\right| \\
&\leq \frac{c_a^{2+\delta}\overline{\mu}_{2+\delta}n^{-(2+\delta)/2}\left(\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{(2+\delta)/2}}{S_n^{2+\delta}}\sum_{i=1}^{n}\left|\zeta_{ni}\right|^{2+\delta}K_{\widetilde{h}\widetilde{\lambda},i}^{2+\delta}(x) \\
&= O\left(\left(n\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{-\delta/2}\right)=o(1),
\end{aligned}
$$

where $\sup_{1\leq i\leq n,n\geq 1}E\left|\varepsilon_{n,i}\right|^{2+\delta}\leq\overline{\mu}_{2+\delta}<\infty$. It follows from Theorems 23.6 and 23.11 of Davidson (1994) that $\widetilde{T}_n\xrightarrow{d}N(0,1)$. The result then follows from the fact that $S_n^2\to c'\Gamma(x)\,c$. Consequently, $N_{n2}\xrightarrow{d}N(0,f^{-2}(x)\,\mathbf{A}^*(x)\,\Gamma(x)\,\mathbf{A}^*(x)')$. This completes the proof. □

Appendix C. Proof of results in Section 4

We first state a lemma that is used to prove the main results in Section 4.

Lemma C.1. *Let $\mathcal{A}_n$ be a real nonstochastic $n\times n$ matrix whose row and column sums are uniformly bounded in absolute value: $\sup_{1\leq j\leq n,n\geq 1}\sum_{i=1}^n\left|\alpha_{n,ij}\right|\leq c_\alpha$ and $\sup_{1\leq i\leq n,n\geq 1}\sum_{j=1}^n\left|\alpha_{n,ij}\right|\leq c_\alpha$ for some $c_\alpha<\infty$. Let $\vartheta_n=n^{-1}U_n'\mathcal{A}_nU_n$ and $\widetilde{\vartheta}_n=n^{-1}\widetilde{U}_n'\mathcal{A}_n\widetilde{U}_n$. Then*

(a) $E[\vartheta_n]=O(1)$, $\operatorname{Var}(\vartheta_n)=O(n^{-1})=o(1)$, *and* $\vartheta_n-E[\vartheta_n]=O_p\left(n^{-1/2}\right)$;
(b) $n^{-1/2}\widetilde{U}_n'\mathcal{A}_n\widetilde{U}_n=n^{-1/2}U_n'\mathcal{A}_nU_n-n^{-1/2}\delta_{\rho n}E[U_n'\overline{G}_{12n}\varepsilon_n]+o_p(1)$, *where* $\delta_{\rho n}\equiv\widetilde{\rho}_n-\rho_n^0$, $\overline{G}_{1n}\equiv G_{1n}'\left(\mathcal{A}_n+\mathcal{A}_n'\right)$ *and* $\overline{G}_{12n}\equiv\overline{G}_{1n}\left(I_n-\gamma_n^0W_{2n}\right)^{-1}$.

Proof of Theorem 4.1. The proof of the consistency of $\widetilde{\gamma}_n$ follows from the same argument as that of Theorem 1 in Kelejian and Prucha (2010). One can readily check the conditions of Lemma 3.1 in Pötscher and Prucha (1997) are fulfilled for our problem. We now establish the asymptotic normality of $\widetilde{\gamma}_n$. Note that

$$
q_n(\gamma_n)=\widetilde{\psi}_n-\widetilde{\Psi}_n\begin{bmatrix}\gamma_n\\ \gamma_n^2\end{bmatrix}=\begin{bmatrix}n^{-1}\widetilde{U}_n'C_{1n}(\gamma_n)\,\widetilde{U}_n\\ n^{-1}\widetilde{U}_n'C_{2n}(\gamma_n)\,\widetilde{U}_n\end{bmatrix},
$$

where $C_{kn}(\gamma_n)\equiv(I_n-\gamma_nW_{2n}')A_{kn}(I_n-\gamma_nW_{2n})$, $k=1,2$. By Assumptions 2, 6 and Fact 1, the row and column sums of $C_{kn}(\gamma_n)$ are uniformly bounded in absolute value. Minimizing Q_n in (4.6) with respect to γ_n yields the first order condition

$$
\left(\frac{\partial q_n(\widetilde{\gamma}_n)}{\partial\gamma_n}\right)'\Upsilon_nq_n(\widetilde{\gamma}_n)=0.
$$

Expanding only $q_n(\widetilde{\gamma}_n)$ about γ_n^0 in the above expression and reorganizing terms yields

$$
\begin{aligned}
&\left(\frac{\partial q_n(\widetilde{\gamma}_n)}{\partial\gamma_n}\right)'\Upsilon_n\frac{\partial q_n(\overline{\gamma}_n)}{\partial\gamma_n}\sqrt{n}\left(\widetilde{\gamma}_n-\gamma_n^0\right) \\
&=-\left(\frac{\partial q_n(\widetilde{\gamma}_n)}{\partial\gamma_n}\right)'\Upsilon_n\sqrt{n}q_n\left(\gamma_n^0\right), \qquad \text{(C.1)}
\end{aligned}
$$

where $\overline{\gamma}_n$ lies between $\widetilde{\gamma}_n$ and γ_n^0 and $\overline{\gamma}_n-\gamma_n^0=o_p(1)$ by the consistency of $\widetilde{\gamma}_n$. Noting that $\frac{\partial q_n(\gamma_n)}{\partial\gamma_n}=-\widetilde{\Psi}_n[1,2\gamma_n]'$, we have

$$
\left(\frac{\partial q_n(\widetilde{\gamma}_n)}{\partial\gamma_n}\right)'\Upsilon_n\frac{\partial q_n(\overline{\gamma}_n)}{\partial\gamma_n}=\begin{bmatrix}1\\ 2\widetilde{\gamma}_n\end{bmatrix}'\widetilde{\Psi}_n'\Upsilon_n\widetilde{\Psi}_n\begin{bmatrix}1\\ 2\overline{\gamma}_n\end{bmatrix}\equiv\widetilde{\Xi}_n.
$$

Let $\Xi_n=\begin{bmatrix}1\\ 2\gamma_n^0\end{bmatrix}'\Psi_n'\Upsilon_n\Psi_n\begin{bmatrix}1\\ 2\gamma_n^0\end{bmatrix}$. The proof is complete if we can show

$$
\widetilde{\Psi}_n=\Psi_n+o_p(1), \qquad \Psi_n=O(1), \qquad \text{(C.2)}
$$

$$
\widetilde{\Xi}_n^{-1}-\Xi_n^{-1}=o_p(1) \quad \text{with with probability approaching 1 as } n\to\infty, \qquad \text{(C.3)}
$$

and

$$
\sqrt{n}q_n\left(\gamma_n^0\right)=\Phi_{n,\gamma\gamma}^{1/2}\xi_n+o_p(1) \quad \text{with } \xi_n\xrightarrow{d}N(0,I_2), \qquad \text{(C.4)}
$$

because then with probability approaching 1 as $n \to \infty$ (w.p.a.1),

$$\sqrt{n}\left(\widetilde{\gamma}_n - \gamma_n^0\right) = \widetilde{\Xi}_n^{-1}\begin{bmatrix}1\\ 2\widetilde{\gamma}_n\end{bmatrix}'\widetilde{\Psi}_n'\Upsilon_n\left[\Phi_{n,\gamma\gamma}^{1/2}\xi_n + o_p(1)\right]$$

$$= \Xi_n^{-1}\begin{bmatrix}1\\ 2\gamma_n^0\end{bmatrix}'\Psi_n'\Upsilon_n\Phi_{n,\gamma\gamma}^{1/2}\xi_n + o_p(1)$$

$$= \left(J_n'\Upsilon_n J_n\right)^{-1}J_n'\Upsilon_n\Phi_{n,\gamma\gamma}^{1/2}\xi_n + o_p(1) \xrightarrow{d} N\left(0, \Omega_\gamma\right). \quad (C.5)$$

First, noting that the elements of $\widetilde{\Psi}_n$ and Ψ_n are of the forms $n^{-1}\widetilde{U}_n'\mathcal{A}_n\widetilde{U}_n$ and $n^{-1}E[U_n'\mathcal{A}_nU_n]$ respectively, we apply Lemma C.1 to obtain (C.2). It then follows from the consistency of $\widetilde{\gamma}_n$ and $\overline{\gamma}_n$ and the Slutsky lemma that $\widetilde{\Xi}_n = \Xi_n + o_p(1)$. By the assumptions on Υ_n and Ψ_n, $\Xi_n \geq \left(1+4(\gamma_n^0)^2\right)\lambda_{\min}\left(\Psi_n'\Psi_n\right)\lambda_{\min}\left(\Upsilon_n\right) \geq c > 0$ for some c, which implies that $0 < \Xi_n^{-1} < \infty$. It follows that w.p.a.1, $\widetilde{\Xi}_n$ is invertible. When this occurs, (C.3) holds.

Next, note that $q_n\left(\gamma_n^0\right) = n^{-1}[\widetilde{U}_n'C_{1n}^0\widetilde{U}_n, \widetilde{U}_n'C_{2n}^0\widetilde{U}_n]'$, where $C_{kn}^0 \equiv C_{kn}\left(\gamma_n^0\right)$ for $k = 1, 2$. By Lemma C.1 and the fact that $U_n = \left(I_n - \gamma_n^0W_{2n}\right)^{-1}\varepsilon_n$,

$$\begin{aligned}n^{-1/2}\widetilde{U}_n'C_{kn}^0\widetilde{U}_n &= n^{-1/2}U_n'C_{kn}^0U_n - n^{-1/2}\delta_{\rho n}E[\varepsilon_n'\overline{C}_{kn}\varepsilon_n]\\ &\quad + o_p(1)\\ &= n^{-1/2}\varepsilon_n'A_{kn}\varepsilon_n - n^{-1/2}\delta_{\rho n}E[\varepsilon_n'\overline{C}_{kn}\varepsilon_n]\\ &\quad + o_p(1),\end{aligned}$$

where for $k = 1, 2$, $\overline{C}_{kn}$ is as defined in (4.9):

$$\begin{aligned}\overline{C}_{kn} &= 2\left(I_n - \gamma_n^0W_{2n}'\right)^{-1}G_{1n}'C_{kn}^0\left(I_n - \gamma_n^0W_{2n}\right)^{-1}\\ &= 2\left(I_n - \gamma_n^0W_{2n}'\right)^{-1}G_{1n}'\left(I_n - \gamma_n^0W_{2n}'\right)A_{kn}. \quad (C.6)\end{aligned}$$

By the proof of Theorem 3.1 with $A_n = \left(I_n - \gamma_n^0W_{2n}\right)^{-1}$, we have $\delta_{\rho n} = \widetilde{\rho}_n - \rho_n^0 = n^{-1}(B'\Omega B)^{-1}B'\Omega Z_n'\left(I_n - \mathbf{S}_{h\gamma}\right)\left(I_n - \gamma_n^0W_{2n}\right)^{-1}\varepsilon_n + o_p\left(n^{-1/2}\right)$. It follows that

$$\sqrt{n}q_n\left(\gamma_n^0\right) = n^{-1/2}\begin{bmatrix}\varepsilon_n'A_{1n}\varepsilon_n + a_{1n}'\varepsilon_n\\ \varepsilon_n'A_{2n}\varepsilon_n + a_{2n}'\varepsilon_n\end{bmatrix} + o_p(1) \quad (C.7)$$

where a_{kn}', $k = 1, 2$, are defined in (4.8). Noticing that the diagonal elements of the matrices $A_{kn}(k = 1, 2)$ are zero, we can apply Theorem A.1 to deduce that the asymptotic variance-covariance matrix of the vector of linear quadratic forms in (C.7) is given by $\Phi_{n,\gamma\gamma} = [\phi_{n\gamma\gamma,kl}]_{k,l=1,2}$, where $\phi_{n\gamma\gamma,kl}$, $k, l = 1, 2$, are defined in (4.10).

Noting that the row and column sums of the matrices A_{kn} are uniformly bounded in absolute value and the elements of a_{kn} are uniformly bounded by a finite constant, we can readily see that $\mathbf{A}_{kn} \equiv A_{kn}$ and $\mathbf{a}_{kn} \equiv a_{kn}$ $(k = 1, 2)$ satisfy the first two conditions of Theorem A.1. By assumption, $\lambda_{\min}\left(\Phi_{n,\gamma\gamma}\right) \geq c_{\Phi_{\gamma\gamma}} > 0$, verifying the third condition of Theorem A.1. Thus, by Theorem A.1, we have

$$\xi_n \equiv \Phi_{n,\gamma\gamma}^{-1/2}n^{-1/2}\begin{bmatrix}\varepsilon_n'A_{1n}\varepsilon_n + a_{1n}'\varepsilon_n\\ \varepsilon_n'A_{2n}\varepsilon_n + a_{2n}'\varepsilon_n\end{bmatrix} \xrightarrow{d} N\left(0, I_2\right). \quad (C.8)$$

By the properties of A_{kn} and a_{kn}, it is straightforward to verify that the elements of $\Phi_{n,\gamma\gamma}$ are uniformly bounded. It follows from (C.7) and (C.8) that $\sqrt{n}q_n\left(\gamma_n^0\right) = \Phi_{n,\gamma\gamma}^{1/2}\xi_n + o_p(1)$. □

Proof of Theorem 4.2. We verify the conditions of Theorem A.1 are met. We have verified in the proof of Theorem 4.1 that the elements of A_{kn} and $a_{kn}(k = 1, 2)$ appearing in v_n satisfy the first two conditions in Theorem A.1. Write

$$\begin{aligned}P_n' &= n^{-1/2}Z_n'\left(I_n - \mathbf{S}_{h\lambda}\right)\left(I_n - \gamma_n^0W_{2n}\right)^{-1}\\ &= n^{-1/2}[p_{n,1\cdot}, \ldots, p_{n,q\cdot}]'.\end{aligned}$$

That is $p_{n,k\cdot}'$ denote the kth row of P_n'. Note that the elements of $p_{n,k\cdot}$ are uniformly bounded by Assumption 5, Lemma A.2, Facts 1 and 2, they satisfy the condition of Theorem A.1 on $\mathbf{a}_{kn} \equiv p_{n,k\cdot}$. By assumption, $\lambda_{\min}\left(\Phi_n\right) \geq c_\Phi > 0$. Thus, by Theorem A.1, we have $\xi_n^a \equiv \Phi_n^{-1/2}\begin{bmatrix}P_n'\varepsilon_n\\ v_n\end{bmatrix} \xrightarrow{d} N\left(0, I_{q+2}\right)$. □

Proof of Theorem 4.3. Let $\widetilde{Y}_n^* \equiv Y_n - \widetilde{\mathbf{m}}\left(X_n\right)$, $Y_n^* \equiv Y_n - \mathbf{m}\left(X_n\right)$, and $\overline{q}_n^*\left(\delta_n\right) \equiv \left(\overline{q}_{1n}^*\left(\delta_n\right), \overline{q}_{2n}^*\left(\delta_n\right)\right)'$, where $\overline{q}_{kn}^*\left(\delta_n\right) = n^{-1}E\left[\left(Y_n^* - \rho_nW_{1n}Y_n\right)'\left(I_n - \gamma_nW_{2n}\right)'A_{kn}\left(I_n - \gamma_nW_{2n}\right)\left(Y_n^* - \rho_nW_{1n}Y_n\right)\right]$ for $k = 1, 2$. Let $\overline{Q}_n^*\left(\delta_n\right) \equiv \overline{q}_n^*\left(\delta_n\right)'\Upsilon_n^*\overline{q}_n^*\left(\delta_n\right)$. By applying Lemma 3.1 in Pötscher and Prucha (1997) with Q_n^* and $\overline{Q}_n^*$ replacing their R_n and $\overline{R}_n$, we can prove the consistency of $\widehat{\delta}_n$. To establish the asymptotic normality of $\widehat{\delta}_n$, we notice that minimizing Q_n^* in (4.12) with respect to δ_n yields the first order condition $\left(\frac{\partial q_n^*(\widehat{\delta}_n)}{\partial\delta_n'}\right)'\Upsilon_n^*q_n^*\left(\widehat{\delta}_n\right) = 0$. By the Taylor expansion for $q_n^*\left(\widehat{\delta}_n\right)$ we have

$$\left(\frac{\partial q_n^*\left(\widehat{\delta}_n\right)}{\partial\delta_n'}\right)'\Upsilon_n\frac{\partial q_n^*\left(\overline{\delta}_n\right)}{\partial\delta_n'}\sqrt{n}\left(\widehat{\delta}_n - \delta_n^0\right)$$

$$= -\left(\frac{\partial q_n^*\left(\widehat{\delta}_n\right)}{\partial\delta_n'}\right)'\Upsilon_n\sqrt{n}q_n^*\left(\delta_n^0\right),$$

where $\overline{\delta}_n$ lies between $\widetilde{\delta}_n$ and δ_n^0 and $\overline{\delta}_n - \delta_n^0 = o_p(1)$ by the consistency of $\widehat{\delta}_n$. First, noting that $\widetilde{Y}_n^* = \widetilde{U}_n + \left(\widetilde{\rho}_n - \rho_n^0\right)W_{1n}Y_n$, we have

$$\begin{aligned}\sqrt{n}q_{kn}^*\left(\delta_n^0\right) &= n^{-1/2}\widetilde{U}_n'\left(I_n - \gamma_n^0W_{2n}\right)'A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)\widetilde{U}_n\\ &\quad + 2n^{-1/2}\left(\widetilde{\rho}_n - \rho_n^0\right)Y_n'W_{1n}'\left(I_n - \gamma_n^0W_{2n}\right)'\\ &\quad \times A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)\widetilde{U}_n\\ &\quad + n^{-1/2}\left(\widetilde{\rho}_n - \rho_n^0\right)^2Y_n'W_{1n}'\left(I_n - \gamma_n^0W_{2n}\right)'\\ &\quad \times A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)W_{1n}Y_n\\ &\equiv \tau_{kn,1} + 2\tau_{kn,2} + \tau_{kn,3}, \quad \text{say.}\end{aligned}$$

By Lemma C.1, $\tau_{kn,1} = n^{-1/2}U_n'\left(I_n - \gamma_n^0W_{2n}\right)'A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)U_n - 2n^{-1/2}\delta_{\rho n}E\{U_n'G_{1n}'(I_n - \gamma_n^0W_{2n})'A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)U_n\} + o_p(1)$. It is straightforward to show that $\tau_{kn,2} = n^{-1/2}\delta_{\rho n}E\{U_n'G_{1n}'\left(I_n - \gamma_n^0W_{2n}\right)'A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)U_n\} + o_p(1)$ and $\tau_{kn,3} = O_p\left(n^{-1/2}\right)$. It follows that $\sqrt{n}q_{kn}\left(\delta_n^0\right) = n^{-1/2}\varepsilon_n'A_{kn}\varepsilon_n + o_p(1)$ and $\sqrt{n}q_n^*\left(\delta_n^0\right) = n^{-1/2}\begin{bmatrix}\varepsilon_n'A_{1n}\varepsilon_n\\ \varepsilon_n'A_{2n}\varepsilon_n\end{bmatrix} + o_p(1) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\Phi_n^*\right)$ by Theorem A.1.

Second, note that

$$\begin{aligned}\frac{\partial q_{kn}^*\left(\widehat{\delta}_n\right)}{\partial\rho_n} &= -2n^{-1}Y_n'W_{1n}'\left(I_n - \widehat{\gamma}_nW_{2n}\right)'A_{kn}\left(I_n - \widehat{\gamma}_nW_{2n}\right)\\ &\quad \times\left(\widetilde{Y}_n^* - \widehat{\rho}_nW_{1n}Y_n\right),\\ &= -2n^{-1}Y_n'W_{1n}'\left(I_n - \gamma_n^0W_{2n}\right)'A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)\widetilde{U}_n + o_p(1)\\ &= -2n^{-1}E\left[U_n'G_{1n}'\left(I_n - \gamma_n^0W_{2n}\right)'A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)U_n\right]\\ &\quad + o_p(1),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial q_{kn}^*\left(\widehat{\delta}_n\right)}{\partial\gamma_n} &= -2n^{-1}\left(\widetilde{Y}_n^* - \widehat{\rho}_nW_{1n}Y_n\right)'W_{2n}'A_{kn}\left(I_n - \widehat{\gamma}_nW_{2n}\right)\\ &\quad \times\left(\widetilde{Y}_n^* - \widehat{\rho}_nW_{1n}Y_n\right)\\ &= -2n^{-1}\widetilde{U}_n'W_{2n}'A_{kn}\left(I_n - \widehat{\gamma}_nW_{2n}\right)\widetilde{U}_n + o_p(1)\\ &= -2n^{-1}E\left[U_n'W_{2n}'A_{kn}\left(I_n - \gamma_n^0W_{2n}\right)U_n\right] + o_p(1).\end{aligned}$$

A similar result holds for $\partial q_{kn}^{*}\left(\widehat{\delta}_n\right)/\partial\delta_n'$. It follows that $\left(\frac{\partial q_n\left(\widehat{\delta}_n\right)}{\partial\delta_n'}\right)' \Upsilon_n^{*}\frac{\partial q_n\left(\widehat{\delta}_n\right)}{\partial\delta_n'} = J_n^{*\prime}\Upsilon_n^{*}J_n^{*} + o_p(1)$, where J_n^{*} is defined in (4.13). By assumption, $\lambda_{\min}\left(J_n^{*\prime}\Upsilon_n^{*}J_n^{*}\right) \geq \lambda_{\min}\left(\Upsilon_n^{*}\right)\lambda_{\min}\left(J_n^{*\prime}J_n^{*}\right) \geq c$ for some $c > 0$ and for sufficiently large n. Consequently,

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\delta}_n - \delta_n^0\right) &= -\left[\left(\frac{\partial q_n^{*}\left(\widehat{\delta}_n\right)}{\partial\delta_n'}\right)' \Upsilon_n^{*}\frac{\partial q_n\left(\widehat{\delta}_n\right)}{\partial\delta_n'}\right]^{-1} \\
&\quad\times\left(\frac{\partial q_n^{*}\left(\widehat{\delta}_n\right)}{\partial\delta_n'}\right)'\Upsilon_n\sqrt{n}q_n^{*}\left(\delta_n^0\right) \\
&= -\left(J_n^{*\prime}\Upsilon_n^{*}J_n^{*}\right)^{-1}J_n^{*\prime}\Upsilon_n^{*}\sqrt{n}q_n\left(\delta_n^0\right) + o_p(1) \xrightarrow{d} N\left(0, \Omega_\delta^{*}\right). \quad\square
\end{aligned}
$$

Appendix D. Proof of results in Section 5

Proof of Theorem 5.1. By assumption, $B_n - B = o_p(1)$, and $\Omega_n - \Omega = o_p(1)$. By the consistency of $\widetilde{\gamma}_n$, (C.2), and the definitions of $\widetilde{J}_n$ and J_n, $\widetilde{J}_n - J_n = o_p(1)$. Let $\widetilde{\Xi}_n = \widetilde{J}_n'\Upsilon_n\widetilde{J}_n$ and $\Xi_n = J_n'\Upsilon_nJ_n$. By the proof of Theorem 4.2 (see (C.3)), w.p.a.1, $\widetilde{\Xi}_n^{-1} - \Xi_n^{-1} = o_p(1)$. To show the consistency of the asymptotic variance-covariance estimator, it remains to show that $\widetilde{\Phi}_n - \Phi_n = o_p(1)$.

Noting that $\Phi_{n,\rho\rho}$ and $\Phi_{n,\rho\gamma}$ involve only one Σ_n (e.g., $\Phi_{n,\rho\rho} = n^{-1}P_n'\Sigma_nP_n$), they can be regarded as the special case of $\Phi_{n,\gamma\gamma}$. So we only prove $\widetilde{\Phi}_{n,\gamma\gamma} - \Phi_{n,\gamma\gamma} = o_p(1)$ since the proof of $\widetilde{\Phi}_{n,\rho\rho} - \Phi_{n,\rho\rho} = o_p(1)$ and $\widetilde{\Phi}_{n,\rho\gamma} - \Phi_{n,\rho\gamma} = o_p(1)$ will be similar and simpler. Note that we can write the (k,l) element of $\Phi_{n,\gamma\gamma}$ as

$$
\begin{aligned}
\phi_{n\gamma\gamma,kl} &= (2n)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_{kln,ij}\sigma_{n,i}^2\sigma_{n,j}^2 \\
&\quad + n^{-1}\boldsymbol{\alpha}_{kn}F_n'\Sigma_nF_n\boldsymbol{\alpha}_{ln} \equiv \phi_{n,kl}^{\perp} + \phi_{n,kl}^{\perp\perp}
\end{aligned}
\tag{D.1}
$$

where $\boldsymbol{\alpha}_{kln,ij} \equiv \left(a_{kn,ij} + a_{kn,ji}\right)\left(a_{ln,ij} + a_{ln,ji}\right)$,

$$
\begin{aligned}
\boldsymbol{\alpha}_{kn} &\equiv -E[\varepsilon_n'\overline{C}_{kn}\varepsilon_n] \\
&= -2E[U_n'\left(I_n - \rho_n^0W_{1n}'\right)^{-1}W_{1n}'\left(I_n - \gamma_n^0W_{2n}'\right)A_{kn}\varepsilon_n],
\end{aligned}
$$

$$F_n' \equiv (B'\Omega B)^{-1}B'\Omega Z_n'\left(I_n - \mathbf{S}_{h\gamma}\right)\left(I_n - \gamma_n^0W_{2n}\right)^{-1}.$$

Similarly, the (k,l) element of $\widetilde{\Phi}_{n,\gamma\gamma}$ is

$$
\begin{aligned}
\widetilde{\phi}_{n\gamma\gamma,kl} &= (2n)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_{kln,ij}\widetilde{\varepsilon}_{n,i}^2\widetilde{\varepsilon}_{n,j}^2 \\
&\quad + n^{-1}\widetilde{\boldsymbol{\alpha}}_{kn}\widetilde{F}_n'\widetilde{\Sigma}_n\widetilde{F}_n\widetilde{\boldsymbol{\alpha}}_{ln} \equiv \widetilde{\phi}_{n,kl}^{\perp} + \widetilde{\phi}_{n,kl}^{\perp\perp}
\end{aligned}
\tag{D.2}
$$

where

$$\widetilde{\boldsymbol{\alpha}}_{kn} \equiv -2\widetilde{U}_n'(I_n - \widetilde{\rho}_nW_{1n}')^{-1}W_{1n}'(I_n - \widetilde{\gamma}_nW_{2n}')A_{kn}\widetilde{\varepsilon}_n,$$
$$\widetilde{F}_n' \equiv (B_n'\Omega_nB_n)^{-1}B_n'\Omega_nZ_n'\left(I_n - \mathbf{S}_{h\gamma}\right)\left(I_n - \widetilde{\gamma}_nW_{2n}\right)^{-1}.$$

By Lemma D.1 below, $\widetilde{\phi}_{n,kl}^{\perp} - \phi_{n,kl}^{\perp} = o_p(1)$. By Lemma D.2 below and the Slutsky lemma, $\widetilde{\phi}_{n,kl}^{\perp\perp} - \phi_{n,kl}^{\perp\perp} = \widetilde{\boldsymbol{\alpha}}_{kn}\widetilde{\boldsymbol{\alpha}}_{ln}(n^{-1}\widetilde{F}_n'\widetilde{\Sigma}_n\widetilde{F}_n) - \boldsymbol{\alpha}_{kn}\boldsymbol{\alpha}_{ln}(n^{-1}F_n'\Sigma_nF_n) = o_p(1)$. Hence $\widetilde{\Phi}_{n,\gamma\gamma} - \Phi_{n,\gamma\gamma} = o_p(1)$. $\square$

Lemma D.1. *Let $\boldsymbol{\sigma}_n^2 = (\sigma_{n,1}^2, \ldots, \sigma_{n,n}^2)'$, $\boldsymbol{\varepsilon}_n^2 = (\varepsilon_{n,1}^2, \ldots, \varepsilon_{n,n}^2)'$, and $\widetilde{\boldsymbol{\varepsilon}}_n^2 = (\widetilde{\varepsilon}_{n,1}^2, \ldots, \widetilde{\varepsilon}_{n,n}^2)'$. Let $\Lambda_n = n^{-1}\left(\boldsymbol{\sigma}_n^2\right)'\mathbf{A}_n\boldsymbol{\sigma}_n^2$, $\overline{\Lambda}_n = n^{-1}(\boldsymbol{\varepsilon}_n^2)'\mathbf{A}_n\boldsymbol{\varepsilon}_n^2$, and $\widetilde{\Lambda}_n = n^{-1}(\widetilde{\boldsymbol{\varepsilon}}_n^2)'\mathbf{A}_n\widetilde{\boldsymbol{\varepsilon}}_n^2$, where $\mathbf{A}_n$ are $n \times n$ real symmetric and nonstochastic matrices. Suppose that the diagonal elements of $\mathbf{A}_n$ are zero and that the row and column sums are uniformly bounded in absolute value by $c_{\mathbf{a}}$. Then*

(a) *$E[\overline{\Lambda}_n] = \Lambda_n = O(1)$, $Var(\overline{\Lambda}_n) = o(1)$, and hence $\overline{\Lambda}_n - \Lambda_n = o_p(1)$;*
(b) *$\widetilde{\Lambda}_n - \overline{\Lambda}_n = o_p(1)$.*

Lemma D.2. *Let $\mathbf{c}_n$ and $\mathbf{d}_n$ be $n \times 1$ vectors whose elements are uniformly bounded in absolute value by c. Let $\widetilde{F}_n$, F_n, $\widetilde{\boldsymbol{\alpha}}_{kn}$, and $\boldsymbol{\alpha}_{kn}$ be as defined in the proof of Theorem 5.1. Recall $\Sigma_n = \mathrm{diag}(\boldsymbol{\sigma}_n^2)$ and $\widetilde{\Sigma}_n = \mathrm{diag}(\widetilde{\boldsymbol{\varepsilon}}_n^2)$, where $\boldsymbol{\sigma}_n^2$ and $\widetilde{\boldsymbol{\varepsilon}}_n^2$ are as defined in Lemma D.1. Then*

(a) *$n^{-1}\mathbf{c}_n'\Sigma_n\mathbf{d}_n = O(1)$, and $n^{-1}\mathbf{c}_n'(\widetilde{\Sigma}_n - \Sigma_n)\mathbf{d}_n = O_p(n^{-1/2}) = o_p(1)$;*
(b) *$n^{-1}F_n'\Sigma_nF_n = O(1)$, and $n^{-1}\widetilde{F}_n'\widetilde{\Sigma}_n\widetilde{F}_n' - n^{-1}F_n'\Sigma_nF_n = o_p(1)$;*
(c) *$\widetilde{\boldsymbol{\alpha}}_{kn} - \boldsymbol{\alpha}_{kn} = o_p(1)$ for $k = 1, 2$.*

Proof of Theorem 5.2. The proof is analogous to and simpler than that of Theorem 5.1 and thus omitted. $\square$

Proof of Theorem 5.3. Let $\overline{\Gamma}_n(x) \equiv n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s\mathbf{Z}_{n,\widetilde{h}}'(x)\,\mathrm{diag}(\mathbf{k}_{\widetilde{h}\widetilde{\lambda}}(x))A_n\overline{\Sigma}_nA_n'\mathrm{diag}(\mathbf{k}_{\widetilde{h}\widetilde{\lambda}}(x))\mathbf{Z}_{n,\widetilde{h}}(x)$, where $\overline{\Sigma}_n = \mathrm{diag}\left(\boldsymbol{\varepsilon}_n^2\right)$ with $\boldsymbol{\varepsilon}_n^2 = (\varepsilon_{n,1}^2, \ldots, \varepsilon_{n,n}^2)'$. By the triangle inequality, $|\widetilde{\Gamma}_n(x) - \Gamma_n(x)| \leq |\Delta_{n1}(x)| + |\Delta_{n2}(x)|$, where $\Delta_{n1}(x) \equiv \overline{\Gamma}_n(x) - \Gamma_n(x)$, and $\Delta_{n2}(x) \equiv \widetilde{\Gamma}_n(x) - \overline{\Gamma}_n(x)$.

First, $\Delta_{n1}(x) = n^{-1}\Pi_{s=1}^{p_c}\widetilde{h}_s\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\boldsymbol{\tau}_{\widetilde{h},i_1}(x)K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\,a_{n,i_1i_2}(\varepsilon_{n,i_2}^2 - \sigma_{n,i_2}^2)a_{n,i_2i_3}K_{\widetilde{h}\widetilde{\lambda},i_3}(x) \times \boldsymbol{\tau}_{\widetilde{h},i_3}(x)$, where $\boldsymbol{\tau}_{\widetilde{h},i}(x)$ is defined in (2.12). Let $\boldsymbol{\tau}_{\widetilde{h},i,k}(x)$ denote the kth element of $\boldsymbol{\tau}_{\widetilde{h},i}(x)$ and $\Delta_{ns,kl}$ the (k,l)th element in $\Delta_{ns}(x)$ $(s = 1, 2)$. Clearly, $E\left[\Delta_{n1,kl}\right] = 0$. Noting that $\boldsymbol{\tau}_{\widetilde{h},i,k}(x)K_{\widetilde{h}\widetilde{\lambda},i}(x)$ is uniformly bounded by $C(\Pi_{s=1}^{p_c}\widetilde{h}_s)^{-1}$ for some constant C by Assumption 4, we have

$$
\begin{aligned}
\mathrm{Var}\left(\Delta_{n1,kl}\right) &= \left(n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s\right)^2\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}\sum_{i_5=1}^{n}\boldsymbol{\tau}_{\widetilde{h},i_1,k}(x) \\
&\quad\times K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\,a_{n,i_1i_2}E(\varepsilon_{n,i_2}^2 - \sigma_{n,i_2}^2)^2 \\
&\quad\times a_{n,i_2i_3}K_{\widetilde{h}\widetilde{\lambda},i_3}(x)\,\boldsymbol{\tau}_{\widetilde{h},i_3,l}(x)\,\boldsymbol{\tau}_{\widetilde{h},i_4,k}(x)\,K_{\widetilde{h}\widetilde{\lambda},i_4}(x) \\
&\quad\times a_{n,i_4i_2}a_{n,i_2i_5}K_{\widetilde{h}\widetilde{\lambda},i_5}(x)\,\boldsymbol{\tau}_{\widetilde{h},i_5,l}(x) \\
&\leq Cn^{-1}\left(\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{-1}n^{-1}\sum_{i_1=1}^{n}K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\,|\boldsymbol{\tau}_{\widetilde{h},i_1,k}(x)| \\
&\quad\times\sum_{i_2=1}^{n}\left|a_{n,i_1i_2}\right|\sum_{i_3=1}^{n}\left|a_{n,i_2i_3}\right| \\
&\quad\times\sum_{i_4=1}^{n}\left|a_{n,i_4i_2}\right|\sum_{i_5=1}^{n}\left|a_{n,i_2i_5}\right| \\
&= O\left(\left(n\prod_{s=1}^{p_c}\widetilde{h}_s\right)^{-1}\right) = o(1).
\end{aligned}
$$

It follows from the Chebyshev inequality that $\Delta_{n1,kl} = o_p(1)$ for $k, l = 1, \ldots, p_c + 1$.

Next, write $\widetilde{m}(x) - m(x) = [\mathbf{s}_{\widetilde{h}\widetilde{\lambda}}(x)'\left(\mathbf{m}(X_n) + U_n\right) + \mathbf{s}_{\widetilde{h}\widetilde{\lambda}}(x)'W_{1n}Y_n(\rho_n^0 - \widetilde{\rho}_n)] - m(x) = d_{n1}(x) + d_{n2}(x) + d_{n3}(x)$, where $d_{n1}(x) \equiv \frac{1}{2}\sum_{i=1}^{n}\mathbf{s}_{\widetilde{h}\widetilde{\lambda}}\left(x_{n,i}, x\right)\left(x_{n,i}^c - x^c\right)'\ddot{m}(x)\left(x_{n,i}^c - x^c\right)\mathbf{1}\left(x_{n,i}^d = x^d\right) + \sum_{i=1}^{n}\mathbf{s}_{\widetilde{h}\widetilde{\lambda}}\left(x_{n,i}, x\right)\left[m\left(x_{n,i}\right) - m(x)\right]\mathbf{1}\left(x_{n,i}^d \neq x^d\right)$, $d_{n2}(x) \equiv \mathbf{s}_{\widetilde{h}\widetilde{\lambda}}(x)'U_n$, and $d_{n3}(x) \equiv -\delta_{\rho n}\widetilde{\mathbf{s}}_{\widetilde{h}\widetilde{\lambda}}(x)'\overline{Y}_n$. By assumption and the proof of Theorem 3.2, $d_{n1}(x) = O(\|\widetilde{h}\|^2 + \|\widetilde{\lambda}\|) = o(n^{-1/4})$ uniformly in x. Let $d_{nj,i} = d_{nj}\left(x_{n,i}\right)$ and $D_{nj} = \left(d_{nj,1}, d_{nj,2}, \ldots, d_{nj,n}\right)'$ for $j = 1, 2, 3$. Then

$$
D_{n2} = \mathbf{S}_{\widetilde{h}\widetilde{\lambda}}U_n, \qquad D_{n3} = -\delta_{\rho n}\mathbf{S}_{\widetilde{h}\widetilde{\lambda}}\overline{Y}_n, \quad\text{and}
\qquad D_n = \sum_{j=1}^{3}D_{nj}.
\tag{D.3}
$$

By (D.3) and the definition of $\widetilde{\varepsilon}_n$, $\widetilde{\varepsilon}_n = (I_n - \widetilde{\gamma}_n W_{2n})\big(U_n - D_n - \delta_{\rho_n}\overline{Y}_n\big) = -(I_n - \widetilde{\gamma}_n W_{2n})D_{n1} + (I_n - \widetilde{\gamma}_n W_{2n})(I_n - \mathbf{S}_{\widetilde{h}\widetilde{\lambda}})U_n - \delta_{\rho_n}(I_n - \widetilde{\gamma}_n W_{2n})(I_n - \mathbf{S}_{\widetilde{h}\widetilde{\lambda}})\overline{Y}_n$. Hence, $\widetilde{\varepsilon}_n - \varepsilon_n = \widetilde{\varepsilon}_n - (I_n - \gamma_n^0 W_{2n})U_n = -(I_n - \widetilde{\gamma}_n W_{2n})D_{n1} - (C_{3n} + \delta_{\gamma n} C_{4n})\varepsilon_n - \delta_{\rho_n}(I_n - \widetilde{\gamma}_n W_{2n})(I_n - \mathbf{S}_{\widetilde{h}\widetilde{\lambda}})\overline{Y}_n \equiv -\eta_{1n} - \eta_{2n} - \eta_{3n}$, where $\delta_{\gamma n} \equiv \widetilde{\gamma}_n - \gamma_n^0$,

$$\begin{aligned} C_{3n} &\equiv (I_n - \gamma_n^0 W_{2n})\mathbf{S}_{\widetilde{h}\widetilde{\lambda}}(I_n - \gamma_n^0 W_{2n})^{-1}, \quad \text{and} \\ C_{4n} &\equiv W_{2n}(I_n - \mathbf{S}_{\widetilde{h}\widetilde{\lambda}})(I_n - \gamma_n^0 W_{2n})^{-1}. \end{aligned} \tag{D.4}$$

It follows that

$$\begin{aligned} \Delta_{n2,kl} &= n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n} \boldsymbol{\tau}_{\widetilde{h},i_1k}(x)\, K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\, a_{n,i_1i_2} \\ &\quad \times (\widetilde{\varepsilon}_{n,i_2}^2 - \varepsilon_{n,i_2}^2) a_{n,i_2i_3} K_{\widetilde{h}\widetilde{\lambda},i_3}(x)\, \boldsymbol{\tau}_{\widetilde{h},i_3l}(x) \\ &= n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n} \boldsymbol{\tau}_{\widetilde{h},i_1k}(x) \\ &\quad \times K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\, a_{n,i_1i_2} a_{n,i_2i_3} K_{\widetilde{h}\widetilde{\lambda},i_3}(x)\, \boldsymbol{\tau}_{\widetilde{h},i_3l}(x) \\ &\quad \times \{\eta_{1n,i_2}^2 + \eta_{2n,i_2}^2 + \eta_{3n,i_2}^2 - 2\eta_{1n,i_2}\varepsilon_{n,i_2} - 2\eta_{2n,i_2}\varepsilon_{n,i_2} \\ &\quad - 2\eta_{3n,i_2}\varepsilon_{n,i_2} + 2\eta_{1n,i_2}\eta_{2n,i_2} + 2\eta_{1n,i_2}\eta_{3n,i_2} \\ &\quad + 2\eta_{2n,i_2}\eta_{3n,i_2}\} \\ &\equiv \sum_{j=1}^{9}\overline{\Delta}_{n2j}, \quad \text{say,} \end{aligned}$$

where the definitions of $\overline{\Delta}_{n2j}$ are self-evident. $\Delta_{n2,kl} = o_p(1)$ provided $\overline{\Delta}_{n2j} = o_p(1)$ for $j = 1, \ldots, 9$. The proof of $\overline{\Delta}_{n2j} = o_p(1)$ is analogous to but simpler than that of $\Delta_{1nj} = o_p(1)$ for $j = 1, \ldots, 9$ in the proof of Lemma D.1(b). So we only sketch the cases of $\overline{\Delta}_{n21}$ and $\overline{\Delta}_{n22}$. Noting that $\sup_i |\eta_{1n,i}| = o_p(n^{-1/4})$, we have $|\overline{\Delta}_{n21}| \le C\sup_i \eta_{1n,i}^2 n^{-1}\sum_{i_1=1}^n |\boldsymbol{\tau}_{\widetilde{h},i_1,k}(x)|\, K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\sum_{i_2=1}^n |a_{n,i_1i_2}| \sum_{i_3=1}^n |a_{n,i_2i_3}| = o_p(n^{-1/2}) = o_p(1)$. Now, by the triangle and Cauchy–Schwarz inequalities,

$$\begin{aligned} E|\overline{\Delta}_{n22}| &\le 2n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n} |\boldsymbol{\tau}_{\widetilde{h},i_1k}(x)| \\ &\quad \times K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\, |a_{n,i_1i_2}||a_{n,i_2i_3}| K_{\widetilde{h}\widetilde{\lambda},i_3}(x)\, |\boldsymbol{\tau}_{\widetilde{h},i_3l}(x)| \\ &\quad \times \sum_{i_4=1}^{n} c_{3n,i_2i_4}^2 \sigma_{n,i_4}^2 \\ &\quad + 2\delta_{\gamma n}^2 n^{-1}\prod_{s=1}^{p_c}\widetilde{h}_s \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n} |\boldsymbol{\tau}_{\widetilde{h},i_1k}(x)| \\ &\quad \times K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\, |a_{n,i_1i_2}||a_{n,i_2i_3}| K_{\widetilde{h}\widetilde{\lambda},i_3}(x)\, |\boldsymbol{\tau}_{\widetilde{h},i_3l}(x)| \\ &\quad \times \sum_{i_4=1}^{n} c_{4n,i_2i_4}^2 \sigma_{n,i_4}^2 \equiv 2(\overline{\Delta}_{n22,a} + \delta_{\gamma n}^2\overline{\Delta}_{n22,b}), \end{aligned}$$

where $c_{sn,ij}$ $(s = 3, 4)$ are the (i, j) elements of C_{sn} defined in (D.4). Observe that $\overline{\Delta}_{n22,a} \le C\overline{\sigma}^2 \sup_{i,j}|c_{3n,i_2i_4}|\, n^{-1}\sum_{i_1=1}^n |\boldsymbol{\tau}_{\widetilde{h},i_1k}(x)|\, K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\sum_{i_2=1}^n |a_{n,i_1i_2}| \sum_{i_3=1}^n |a_{n,i_2i_3}| \sum_{i_4=1}^n |c_{3n,i_2i_4}| = O((n\Pi_{s=1}^{p_c}\widetilde{h}_s)^{-1})$, and similarly $\overline{\Delta}_{n22,b} \le C\overline{\sigma}^2\sup_{i,j}|c_{4n,i_2i_4}|\, n^{-1}\sum_{i_1=1}^n |\boldsymbol{\tau}_{\widetilde{h},i_1k}(x)|\, K_{\widetilde{h}\widetilde{\lambda},i_1}(x)\sum_{i_2=1}^n |a_{n,i_1i_2}|\sum_{i_3=1}^n |a_{n,i_2i_3}|\sum_{i_4=1}^n |c_{4n,i_2i_4}| = O(1)$. It follows from the Markov inequality that $\overline{\Delta}_{n22} = O_p((n\Pi_{s=1}^{p_c}\widetilde{h}_s)^{-1}) + O_p(n^{-1}) = o_p(1)$. Analogously, we can show that $\overline{\Delta}_{n2j} = o_p(1)$ for $j = 3, \ldots, 9$. Hence $\Delta_{n2} = o_p(1)$ and

$$\begin{aligned} \|\widetilde{\Gamma}_n(x) - \Gamma_n(x)\| &\le \|\Delta_{n1}(x)\| + \|\Delta_{n2}(x)\| = o_p(1) + o_p(1) \\ &= o_p(1). \end{aligned} \tag{D.5}$$

The consistency of $\widetilde{f}_n(x)$ and $\widetilde{\varphi}_n(x)$ follows from Assumptions 3 and 4, which together with (D.5), implies $\widetilde{\Omega}_{n,M}(x) - \Omega_{n,M}(x) = o_p(1)$ by the Slutsky lemma. □

References

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71, 1795–1843.

Aitchison, J., Aitken, C.G.G., 1976. Multivariate binary discrimination by the kernel method. Biometrika 63, 413–420.

Baltagi, B.H., Li, D., 2001. LM tests for functional form and spatial correlation. International Regional Science Review 24, 194–225.

Basile, R., Gress, B., 2005. Semi-parametric spatial auto-covariance models of regional growth behavior in Europe. Région et Développement 21, 97–118.

Cai, Z., Li, Q., 2008. Nonparametric estimation of varying coefficient dynamic panel data models. Econometric Theory 24, 1321–1342.

Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford University Press.

Fan, J., 1992. Design-adaptive nonparametric regression. Journal of the American Statistical Association 87, 998–1004.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman and Hall.

Gress, B., 2004. Using semi-parametric spatial autocorrelation models to improve hedonic housing price prediction. Mimeo. Dept. of Economics, UC, Riverside.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Henderson, D.J., Carroll, R.J., Li, Q., 2008. Nonparametric estimation and testing of fixed effects panel data models. Journal of Econometrics 144, 257–275.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 377–398.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.

Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.

Lee, L-F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252–277.

Lee, L-F., 2004. Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lin, X., Lee, L-F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157, 34–52.

Linton, O., 1995. Second order approximation in the partially linear regression model. Econometrica 63, 1079–1112.

Li, Q., Racine, J.S., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press.

Liu, X., Lee, L., Bollinger, C.R., 2010. An efficient GMM estimator of spatial autoregressive models. Journal of Econometrics 159, 303–319.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency rates. Journal of Time Series Analysis 17, 571–599.

Newey, W.K., 1990. Efficient instrumental variables estimation of nonlinear models. Econometrica 58, 809–837.

Pace, P.K., Barry, R., Slawson Jr., V.C., Sirmans, C.F., 2004. Simultaneous spatial and functional form transformation. In: Anselin, L., Florax, R., Rey, S.J. (Eds.), Advances in Spatial Econometrics. Springer-Verlag, Berlin, pp. 197–224.

Paelinck, J.H.P., Klaassen, L.H., 1979. Spatial Econometrics. Gower Press, Aldershot.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press, Cambridge.

Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models: Asymptotic Theory. Springer Verlag, New York.

Racine, J., Li, Q., 2004. Nonparametric estimation of regression function with both categorical and continuous data. Journal of Econometrics 119, 99–130.

Robinson, P.M., 1988. Root-n consistent semiparametric regression. Econometrica 56, 931–954.

Su, L., Jin, S., 2010. Profile quasi-maximum likelihood estimation of spatial autoregressive models. Journal of Econometrics 157, 18–33.

Su, L., Ullah, A., 2006. Profile likelihood estimation of partially linear panel data models with fixed effects. Economics Letters 92, 75–81.

Su, L., Yang, Z., 2009. Instrumental variable quantile estimation of spatial autoregressive models. Working paper. Singapore Management University..

van Gastel, R.A.J.J., Paelinck, J.H.P., 1995. Computation of Box–cox transform parameters: a new method and its application to spatial econometrics. In: Anselin, L., Florax, R.J.G.M. (Eds.), New Directions in Spatial Econometrics. Springer-Verlag, pp. 136–155.

Xiao, Z., Linton, O.B., Carroll, R.J., Mammen, E., 2003. More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. Journal of the American Statistical Association 98, 980–992.

Yang, Z., Li, C., Tse, Y.K., 2006. Functional form and spatial dependence in dynamic panels. Economics Letters 91, 138–145.